University of Northern Colorado
Scholarship & Creative Works @ Digital UNC

Dissertations | Student Research

5-1-2014

Study of statistical efficiency on the effects of non-compliant reporting and item list size from the indirect questioning techniques: random response and non-random response models

Caroline Bublitz Emsermann

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

A STUDY OF STATISTICAL EFFICIENCY ON THE EFFECTS OF NON-COMPLIANT REPORTING AND ITEM LIST SIZE FROM THE INDIRECT QUESTIONING TECHNIQUES: RANDOM RESPONSE AND NON-RANDOM RESPONSE MODELS

A Dissertation Submitted in Partial Fulfillment
of the Requirements of the Degree of
Doctor of Philosophy

Caroline Bublitz Emsermann

College of Education and Behavioral Sciences
Program of Applied Statistics and Research Methods

May 2014

This Dissertation by: Caroline Bublitz Emsermann

Entitled: *A Study of Statistical Efficiency on the Effects of Non-compliant Reporting and Item List Size from the Indirect Questioning Techniques: Random Response and Non-random Response Models*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in College of Education and Behavioral Sciences, Program of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

____

Jay Schaffer, Ph.D., Research Advisor

____

Trent L. Lalonde, Ph.D., Committee Member

____

Khalil Shafie, Ph.D., Committee Member

____

Robert Heiny, Ph.D., Faculty Representative

Date of Dissertation Defense ____

Accepted by the Graduate School

____

Linda L. Black, Ed.D., LPC
Dean of the Graduate School and International Admissions

ABSTRACT

Bublitz Emsermann, Caroline. *A Study of Statistical Efficiency on the Effects of Non-compliant Reporting and Item List Size from the Indirect Questioning Techniques: Random Response and Non-random Response Models*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2014.

Estimating prevalent rates of sensitive behaviors using self-report measures generally resulted in bias estimates when direct questioning approaches were used. Random Response (RR) and Non-random Response (NRR) models were developed to provide an additional layer of confidentiality that was meant to illicit more honest reporting. Despite these efforts, there was evidence that survey participants using these techniques do not always report honestly, and as a result, estimates from these techniques were biased. The current study examined the statistical efficiency, using the ratio of MSE between the RR models, the unrelated question technique (UQT) and forced choice technique (FCT) and the NRR models, the item count technique (ICT), double item count technique (DICT) and the single sample count technique (SSC). Simulations of a large range of sensitive prevalent rates and sample sizes were performed where estimates were compared in terms of increasing levels of non-compliance. In addition, for NRR models exclusively, techniques were compared similarly by list size (3-item, 4-item, and 5-item) as well as between each of the NRR models (ICT, DICT, and SSC). Results of the study indicated that the UQT optimal model was the most efficient of the techniques in the presence of equivalent non-compliance rates. However, if the DICT optimal 5-item model improved compliance, this model became more efficient depending on the sensitive prevalent rate estimated and the sample size. As a result, the study demonstrated that in certain situations, the non-random response double item count technique optimal model was as or more efficient than the random response unrelated questioning technique optimal model. Applications of the findings and the development of general guidelines were discussed.

ACKNOWLEDGEMENTS

When my husband, Markus, wrote his acknowledgements for his Ph.D. dissertation, he started out with a quote from Paul McCartney, “it has been a long and winding road.” I now understand exactly what he meant. Pursuing a Ph.D. is a commitment that takes a great deal of patience and a lot of effort. I would never have seen it through without the support and encouragement from so many people. First, I would like to thank my advisor, Dr. Jay Schaffer, for providing me with this interesting topic as well as his guidance throughout my time as a Ph.D. student--he is one of the most talented instructors and I thoroughly enjoyed every one of his courses. Also thanks to my committee members, Dr. Trent Lalonde, Dr. Khalil Shafie, and Dr. Robert Heiny, for their careful review, comments, and support, especially since they were forced to review my dissertation over their Thanksgiving holiday! To Drs. Lalonde and Shafie, I too enjoyed your instruction, especially in linear models and categorical data analysis--even the difficult test, homework, and comprehensive exam questions--which prepared me for real world applications. I also would like to thank the heart of this department, Keyleigh Gurney, without whom the department could not exist. I cannot thank her enough for everything she did for me--from scanning in homework assignments to helping me sign up for courses when I missed deadlines! Also for waiting with me to hear whether I had passed my comprehensive oral exam, dissertation proposal, and defense. You are one in a million!

Thanks also go out to the faculty at the Department of Family Medicine, especially the support of Wilson Pace, Perry Dickinson, and David West. Thanks for letting me attend meetings as conference calls from Greeley as well as rearranging deadlines and meetings just to fit my school schedule. I am proud to be part of the Department and look forward to my continued work with each of you. Very special and heartfelt thanks also go out to my mentors, Dr. Karen Kafadar and Dr. Miriam Dickinson, two statisticians whose encouragement and faith saw me through. I can only hope one day I am the statistician you both are--you are such inspirations. I would also like to thank the many friends and UNC colleagues who supported me throughout including, Marian Craig, Deb, Cheryl, Christina, Jenny, Lura, Stephanie, Karen, Gerry, Bill, Leytonia, Stephanie M, Jamie, Kimberly, Aman, Maryann, Kevin, Intesar, and Jamil--I could not have made it through the program without such good company.

Special thanks also goes out posthumously to Rose Kelley, my fifth grade teacher whose encouragement gave me the courage to live my dream--and who also gave me my first mathematical task when she challenged me in a race to write out the 12 times tables on the chalkboard! Kim Saltus, who inspired me to always live life to the fullest and Bill (Buddha) Owens--for showing that life is about love, laughter, and good times with friends. I miss you all so much!

A special thanks also goes out to all the special pets (both past and present) who make life so enjoyable, including my Deutsch Morsel, Bernie, TJ, JR, Dino, Otack, Kung Fu, Avie, Spinoza, Katia, Dana, Dobie, Bailey, Harlequin, Ajax, Barnie, Zulu, and my special horses Lucy (I am coming your way), Notie, Rooster, Bear and Petey--with a very special thanks to my Cogito and Otis, who both started out with me on this journey but were not able to see it through--you will live in my heart forever. Also a big thank you to my favorite musicians who entertained and kept me company with their music on the long drives to and from Greeley: Paul McCartney (for taking me to all the great cities), Elton John, Keith Emerson and Greg Lake (thanks for the good time in PA . . . you are both awesome!), Queen, Bob Seger, and Led Zeppelin.

I would also like to thank my family, especially my sister Lish--who has taught me how important it is to live one day at a time, Susan--for being the proud big sister (not to mention the books on tapes she gave me for the long commute), and my parents, who inspired, supported, and encouraged me to always live my dream! Also, a special thanks to my niece and nephews, Jimmy, Stephen, Richard, Willy, Max, Gretchen, Adam, and Milo--you make me proud of the next generation.

My most heartfelt thanks, however, goes to my husband, soul mate, and best friend, Markus. Your patience and endurance during this time was vital to my success. I thank you for your honesty, love, support, mathematical genius, and all the times you cleaned the kitchen and made dinner! I love you so much and am looking forward to our future adventures together.

Lastly, I must thank my 1998 Honda Civic--for transporting me between Parker and Greeley three times a week! Thank you for always getting me there safely and without incident!

The adventure as a Ph.D. student is now ending and a new life of freedom begins. I am reminded of a quote from Bernie Taupin's lyric "Live Like Horses," which sums up my feelings best: "I've spent too long in the belly of the beast and now I shall be free . . . break out the stalls and (one day) we'll live like horses." Well, today is that day!

TABLE OF CONTENTS

CHAPTER
IV. continued

APPENDICES

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Table

CHAPTER I

INTRODUCTION

Background

Self-reports in surveys have become increasingly relied upon to measure attributes of sensitive behaviors (Fox & Tracy, 1986). During the past three decades, survey instruments have been used to collect information and provide estimation for such sensitive topics as illicit drug use, the use of performance enhancing drugs (PED) among athletics, sexual behaviors, income and voting behaviors (Tourangeau & Yan, 2007). Surveys are the primary source of data collection of sensitive topics since actual data are either nonexistent (such as unreported victimization experiences among prisoners) or confidential (such as accessing patient's medical records; Fox & Tracy, 1986). Since the administration of surveys rely entirely on self-report, due to the sensitive nature of these areas, estimates from questions eliciting information on sensitive topics were often misreported resulting in biased estimates that can be misleading (Fox & Tracy, 1986; Greenberg, Abul-Ela, Simmons, & Horvitz, 1969; Tourangeau & Yan, 2007; Warner, 1965).

The reason for such biases was due to the fact that human subjects were selected to represent target populations where estimation is based on self-report instead of actual measurements. Estimation by survey is subject to two additional sources of variation. The first is sampling error which is the variation in the data due to the fact that a sample of subjects were selected. Sampling error can be minimized prior to data collection if the sampling design or selection of subjects from the target population was made to be representative (Thompson, 2002). In addition, complex estimation procedures have been developed that minimize sampling error and improve efficiency of survey estimates (Fox & Tracy, 1986; Thompson, 2002). A more problematic source of variation, especially in the case of sensitive topics, was non-sampling error or systematic error. Thompson (2002) discussed several sources of non-sampling error including non-response, data entry error, or detectability problems. In general these sources could be categorized into two types of non-sampling error: random error which cancels out over repeated measures (data entry errors) and non-random error that does not cancel out over repeated samples (non-response). In the case of sensitive topics, Greenberg et al. (1969) discussed two particular problematic sources of non-random error that resulted in bias estimates:

1. Non-response or the refusal to answer the sensitive question.
2. Deliberate falsification of information or answering the sensitive question dishonestly.

In the former case estimates were biased since a subset of survey participants choose not to respond to the sensitive question. This subset of non-respondents was generally not typical of the segment of the population as a whole. The sample therefore was unrepresentative of the population resulting in a bias estimate (Thompson, 2002). In addition to the bias, the variation of the estimate increased since the number of participants answering the question declined resulting in a less efficient statistic. In the latter case of deliberate falsification, validity of the survey instrument was undermined since estimates were distorted. Validity is defined as the extent to which the survey instrument measures what it purports to measure, in this case the prevalence of the sensitive behavior. As a result, deliberate falsification also resulted in estimation bias that either over estimated the prevalence or under estimated the prevalence. Unlike random error which canceled out over repeated measurements, response bias remained intact and the distortion between the true and reported response persisted (Fox & Tracy, 1986). In the case of sensitive topics, response biases due to falsification are well documented in the literature (Belli, Traugott, & Beckmann, 2001; Fu, Darroch, Henshaw, & Kolb, 1998; Tourangeau & Yan, 2007)

Tourangeau, Rips, and Rasinki (2000) classified three distinct reasons for these types of reporting biases: intrusiveness, threat of disclosure and socially desirable responding. Intrusiveness was where the participant feels as if their privacy has been violated. Examples include questions regarding personal income or religious beliefs. Threat of disclosure refers to participants who feel their confidentiality in disclosing sensitive information was not guaranteed. Even with the assurance of non-disclosure, participants may be hesitant to answer the sensitive question honestly. For instance, an employee could feel reluctant to truthfully respond to questions regarding marijuana use if the confidential survey were administered at their place of employment. Lastly, socially desirable responding was where participants respond to questions in a manner that conforms to socially acceptable behavior or norms. For instance a non-voter who claims to have voted in the last election may do so because of concerns that a truthful response would be seen as socially unacceptable. Socially desirable responding was especially influential and can lead to response bias that underestimates or overestimates prevalence rates depending on the behavior analyzed. Behaviors seen as socially undesirable such as illicit drug use, abortion rates, or use of PED among athletes were often underestimated. In turn, behaviors seen as socially acceptable such as voting, seat belt usage, or exercising were often overestimated (Jann, Jerke, & Jrumpal, 2012).

In an attempt to reduce these types of reporting biases, in 1965, Stanley L. Warner introduced a model designed to elicit indirect responses to sensitive questions. The random response (RR) model was designed to reduce systematic biases which Warner (1965) referred to as "evasive answer bias". Warner argued that if survey participants felt that their response could not be directly linked to the sensitive behavior, they would answer more truthfully resulting in a decrease of evasive answer bias. Warner came up with an ingenuous method where the survey participant would respond to one of two complementary questions with a known probability:

1. You are a member of population A (i.e., π),
2. You are not a member of population A, (i.e., 1- π)

where π is the proportion of those with the sensitive attribute. Since the two questions referred to complementary populations, the participant was either a member of one or the other. In Warner's original model, a spinner was used which would select the sensitive question with probability, p and its complement with probability (1-p). Since then, randomized devices have included other mechanisms with known probabilities such as flipping coins, selecting cards, and rolling dice. The survey participant controls the device that was concealed from the interviewer, who has no knowledge of the question selected, only of the response given. After using the device to select the question, the participant responded either "yes" or "no". The interviewer who does not know which question was selected simply tracks the number of "yes" and "no" responses. Since the probabilities of selecting each question was known, prevalence of the sensitive behavior was estimated without directly linking the observed response and the variable of interest to the individual. As a result, in theory, the participant felt more compelled to answer honestly compared to direct questioning methods.

Because of the randomized device, additional variation must be accounted for in the model. The variance of Warner's estimate can be decomposed into two portions: the variation due to sampling (i.e., binomial variance) and the additional variation due to the randomized device. This additional variation was the cost associated with eliciting more honest responses compared to direct questioning. In fact, there was a direct correlation between the burden of cooperation of the survey participant and the randomized device or p (the probability the sensitive question is selected). If the probability of selection is close to 1, as Warner (1965) pointed out, the burden of cooperation falls to the survey respondent who was more likely to select the sensitive question and therefore feel exposed. For instance, if $p = 1$, the burden of cooperation fell entirely on the survey participant who in effect was asked directly if they were a member of the sensitive group. As p moves from 1 to 0.50, however, the burden of cooperation shifted to the interviewer since the survey participant was now less likely to select the sensitive question and therefore less exposed. In this case, the respondent provides useful but not absolute information regarding their sensitive behavior status. According to Warner, a p less than unity was expected to increase cooperation and at the same time reduced the evasive answering. Due to the additional variation, randomized response techniques were more beneficial than techniques of direct questioning if the trade off between additional variation and bias was warranted. In his original paper, Warner (1965) demonstrated that his randomized response technique produced more efficient estimates compared with the direct questioning technique when evasive answer bias was high. By comparing the mean squared errors for several values of *p*, a fixed sample size and the sensitive population prevalence between 0.50 and 0.60, Warner showed that as evasive answer bias and *p* increased so did the efficiency of the randomized response technique compared to estimates that used direct questioning (Warner, 1965). However, Warner's efficiency study was limited in several aspects. The study assumed that participants always responded truthfully when the randomized response model was used but not when the self-direct technique was applied. As a result, statistical efficiency was not addressed in the case of biased estimates from the random response model. Secondly, the proportion of the sensitive behavior used in the simulations studies were larger than prevalence rates of more sensitive topics, such as drug use, or abortion rates. Lastly, since both questions in Warner's model referred to the sensitive behavior, did the model truly reduce evasive response bias?

In certain situations, Greenberg et al. (1969) showed that Warner's technique produced less efficient estimates compared with the conventional method of direct questioning. By selecting smaller and more appropriate proportions of the sensitive attribute (between 10% and 20%), unless a substantial amount of evasive answer bias was present, Greenberg et al. (1969) demonstrated that estimates from Warner's model were generally less efficient compared to the direct questioning technique. Greenberg et al. (1969) also noted the inefficiency of Warner's method when *p* is closer to 0.50, which is the recommended level if complete cooperation were to be made. In addition, Greenberg et al. (1969) pointed out that both randomly selected questions in the model related directly to the sensitive behavior since each referred to the complement population of the other. This could raise suspicions if the respondent feels as if a "mathematical trick" is at play and potentially result in a higher than expected rate of evasive answering (Fox & Tracy, 1986). In an attempt to reduce these limitations, Greenberg et al. (1969) provided the theoretical framework of the Random Response (RR) technique introduced by Walt R. Simmons (Horvitz, Shah, & Simmons, 1967), the unrelated question technique (UQT). Instead of using two complementary questions, Simmons' technique posed two unrelated questions that were selected in the same manner - via a randomized device with known probabilities:

1. You are a member of population A, (i.e., π_s)
2. You are a member of population Y, (i.e., π_{ns})

where population π_s, the proportion of the population with the sensitive attribute and population π_{ns}, the proportion of the population with the innocuous attribute, were unrelated or in statistical terms, uncorrelated. For instance a researcher using the unrelated question technique to estimate prevalence among elite athletes utilizing PEDs could have participants use a randomized device to select one of the two unrelated questions: "During the past 12 months did you ever use any performance enhancing drugs (PEDs)?" or "Do you have a subscription to a local newspaper?" By posing a question that was completely independent of the sensitive question, participants were made to feel more secure since the population referred to in the innocuous question was not tied to the sensitive population. As a result, no direct connection would be made between the participant and the sensitive behavior.

In contrast to Warner's model where one unknown population parameter was estimated, in the case of the unrelated question model, two unknown population parameters were estimated: the prevalence of the sensitive behavior, π_s and the prevalence of the innocuous behavior, π_{ns}. When the prevalence of the innocuous behavior was unknown, the technique requires two separate samples, where participants from each were provided with the same survey questions, only the probability of selecting the sensitive question in the first sample, p_1, is the complement of the probability of selecting the sensitive question in the second sample, $p_2 = 1 - p_1$. After selecting the appropriate parameters, under the assumption of truthful responses, both samples were used to calculate unbiased estimates of the sensitive and innocuous behavior. Greenberg provided an optimal method in which the selection of the model parameters p_1, p_2, π_{ns}, n_1, n_2 could be made in such a manner as to attain a variance as close to the binomial as possible. Although selecting the parameters in this fashion was meant to minimize variation, compromising the cooperation of the participant should be the researcher's primary concern. Model parameters should therefore be selected in a manner that maximizes both confidentiality and statistical efficiency.

In order to examine the gain (or loss) caused by the unrelated attribute, Greenberg et al. (1969) compared the efficiency of the unrelated technique to Warner's technique. Unlike Warner, Greenberg et al. (1969) considered increasing levels of evasive answer bias, while holding all parameters in both models fixed. Prevalence rates of the sensitive attribute and innocuous attribute were kept relatively low ($\pi_s = 0.20$, $\pi_{ns} = 0.10$) in accordance to the selection criteria described previously. By comparing the mean squared error of both models, Greenberg et al. (1969) determined that the unrelated question technique was superior to Warner's technique when the rate of truthful responses was greater than or equal to the corresponding rate in the Warner technique. As a result, Greenberg et al. (1969) demonstrated that if the unrelated question model improved even very slightly the probability of reporting truthfully a respondent's membership in the sensitive behavior group, the mean squared error showed gains out of proportion to such increases in truthfulness.

Efficiency, however, was improved if the prevalence of the innocuous behavior were actually known. According to Horvitz et al. (1967), if population parameters of the innocuous behavior were incorporated into the randomized device, the innocuous question would be eliminated entirely and a single sample used. As a result, the cost associated with estimating the innocuous behavior would be eliminated and a more efficient estimate of the sensitive behavior gained. This random response (RR) method became known as the forced response or forced choice technique (FCT) and was further developed by Boruch (1971). The technique used the random device to "force" the survey participant to answer the sensitive question in a specific way. An example of the technique would be using the sum of two dice to determine how the survey participant responds. For instance, if the sum of two dice were 2, 3, or 4, the participant would be instructed to respond "yes", if instead the sum were between 5 and 10, the participant would be instructed to respond truthfully to the sensitive question and lastly if the sum were 11 or 12, the participant would be instructed to respond "no". Since the associated probabilities of each outcome were known (i.e., 1/6, 3/4, and 1/12), maximum likelihood estimates of the sensitive behavior were derived. Greenberg et al. (1969) showed that for fixed prevalence of the sensitive behavior at the very low rates of $\pi_s = 0.20$, and $\pi_s = 0.05$, estimates were more efficient compared to estimates from both the unrelated question model where the prevalence of the innocuous question was unknown and Warner's model.

Since their introduction, random response (RR) techniques have been further developed and improved (Kuk, 1990; Pal & Singh, 2012; Saha, 2010). In particular a great deal of effort has been made in increasing the efficiency of estimates (Lensvelt-Mulders, Hox, & van der Heijden, 2005; Pal & Singh, 2012; Saha, 2010) as well as improving psychological features of the technique in such a way as to encourage more participation (Lensvelt-Mulders, Hox, & van der Heijden, 2005; Lensvelt-Mulders, Hox, van der Heijden, & Maas., 2005) performed an extensive meta-analysis on a variety of RR techniques and sensitive topics. The authors determined that random response techniques produced higher prevalence estimates of sensitive attributes compared with conventional methods such as direct questioning. In addition, response rates were generally higher when randomized techniques are used (Clark& Desharnais, 1998; Fox & Tracy, 1986).There were limiting factors to the technique, however. For one, there was the potential of point estimates falling out of the (0, 1) range. This was especially apparent in the unrelated question technique when model parameters were not selected properly (Greenberg et al., 1969). Another issue occurred when participants did not follow the instructions and responded in a manner that was not truthful. This type of responding was referenced in the literature by several terms including "cheaters" (Clark & Desharnais, 1998), non-compliance (van den Hout & Klugkist, 2009), and "self-protected" (SP) response (Böckenholt & van der Heijden, 2004). In each case, the definition was similar, whether intentional or unintentional survey participants fail to follow instructions of the random response (RR) model resulting in misreporting. In the case of a SP response, the definition think you want to leave this capitalized--not hyphenated was extended somewhat by assuming that the response given by the participant provides no information. No matter what the selected question, whether the sensitive or innocuous or a forced response, the participant responded in the negative. In either case, when a substantial number of survey participants responded in these manners, the probabilities associated with the responses were no longer known. Prevalence estimates were therefore distorted which resulted in a biased estimate - defeating the primary purpose of using the RR technique in the first place.

A review of the literature revealed several occurrences where estimates of sensitive behaviors from RR techniques under estimated prevalence of the behavior compared with more accurate data sources. Studies included underestimating abortion rates (Shimizu & Bonham, 1978), racial prejudice, political, and moral issues (Wiseman, Moriarty, & Schafer, 1975) and more recently the use of Mephedrone (Petróczi et al., 2011). In each case non-compliance was cited as a potential biasing factor. In fact Wiseman et al. (1975) included a supplementary question asking participants if they were confident that the random device protected their anonymity. Since 20% of the RR participants felt the interviewer knew what question they selected, Wiseman et al. (1975) concluded that distrust in the randomized technique could have biased the results.

Since RR models were not immune to “cheaters” or participants who are “non-compliant” or are “self protected” (SP), estimates from RR models were subject to distortions. For the remainder of this project, the study chooses the terms non-compliant or SP response in reference to survey participants who do not follow instructions of the randomized response technique. A review of the literature indicated that the primary focus of non-compliance in random response (RR) models has been the development of methods that estimate or adjust for the non-compliance resulting in an adjusted estimate of the sensitive behavior (Böckenholt, Barlas, & van der Heijden, 2009; Böckenholt & van der Heijden, 2004, 2007; Clark & Desharnais, 1998; Cruyff, van den Hout, van der Heijden, & Böckenholt, 2007; Ostapczuk, Moshagen, Musch, & Zhao, 2010; van den Hout, Böckenholt, & van der Heijden, 2010; van den Hout & Klugkist, 2009;). In all cases, no assumptions were made regarding the intentions of the participants who was non-compliant and with the exception of Böckenholt and Van der Hejiden (2007), non-compliance was estimated in terms of the self-protected “no” response--where participants provide no information about either question.

Other methods were developed using several surveys conducted in the Netherlands on social security regulation infringements that included a series of randomized response questions--using a variety of RR techniques including Kuk’s (1990) method and the forced response method--to estimate specific sensitive behaviors that included social security regulation infringements and social welfare fraud. Surveys were fielded in 2000, 2002, and 2004 and included a series of questions regarding the specific sensitive behaviors (i.e., social security fraud) that were ordered from less to more severe violations. Estimation of non-compliance included using item-response models that incorporated a person level estimate (Böckenholt & van der Heijden, 2004, 2007), mixture components (Böckenholt & van der Heljden, 2007) and log linear models (Böckenholt & van de Hejden, 2007) using semi-parametric item response models following directly from latent class models (Böckenholt et al., 2009). Bayesian techniques were also used to estimate non-compliance (van den Hout & Klugkist, 2009). As a result, the primary focus was to develop methods that provide adjusted prevalence estimates of both the sensitive attribute and non-compliance. In general, these methods produced higher prevalent estimates of the sensitive attribute and better fit statistics than the models that do not adjust for non-compliance. With the exception of Greenberg et al. (1969) and Clark and Densharnais (1998), there were no extensive studies that compared the effects of non-compliance on estimates between RR techniques, such as the unrelated question technique (UQT) or the forced-choice technique (FCT). In particular, how different rates of non-compliance distorted estimates of differing sensitive behavior prevalence. In addition, although efficiency was studied in several empirical and simulation studies, there appeared to be no extensive research examining the efficiency between RR techniques in the presence of non-compliant distortions.

Despite the effort to produce unbiased estimates of sensitive topics, RR models had several limitations. For one, they include an additional source of variation since a random device was used. Because of this, in order to improve efficiency, models generally required larger sample sizes compared to models that use a direct questioning approach. Secondly, since many of the techniques required a random device such as spinners, coins, and cards, additional costs were associated with the model. In addition, non-compliant responses distorted estimates despite the efforts of the randomized device to offer additional protection since survey participants were often forced to respond positively to behaviors they had never engaged or to directly respond to a sensitive question.

A second class of indirect response techniques, termed non-randomized response (NRR) models, improved on several of these limitations by eliminating the need of a randomized device and utilizing an even more evasive response method where survey participants indirectly responded to the sensitive question. Unlike the random response (RR) models, participants were asked to respond to a series or combination of questions in which they simply record the number of questions in the series or select a combination of questions for which they agree. This eliminated the need for the randomized device. The item count technique (ICT) was probably the most widely used of the non-random response (NRR) techniques. Also referred to as the list technique or the unmatched count technique (UCT), it was empirically demonstrated in a study by Dalton, Wimbush, and Daily (1994) that investigated illicit workplace behaviors of auctioneers, but was first introduced by Miller (1984). Since its introduction the technique has become widely used and referenced in the literature (Cobb, 2001; Dalton et al., 1994; Dalton, Daily, & Wimbush, 1997; Droitcour et al., 1991; Kuklinski & Cobb 1998; Kuklinski, Cobb, & Gilens, 1997; Kuklinski, Sniderman, et al., 1997; Miller, 1984; Miller, Cisin, & Harrel, 1986; Sniderman & Grob 1996; Tsuchiya, Hirai, & Ono, 2007). The technique used an item list of questions that included a series of innocuous questions and the sensitive question. The innocuous questions could include, “I subscribe to a newspaper,” “I have resided in two or more states” and the sensitive question could be “I have cheated on my income taxes.” Subjects for the study were selected and randomly assigned to two groups. The first group of subjects was given a survey with an item list that only included the innocuous questions. They were then instructed to read each question and report the number of questions for which they agreed (i.e., number of “yes” responses). Since only the aggregate of yes responses were required, there was no direct link between the actual questions for which agreement was made and the participant. The second group of subjects was given the same item list of innocuous questions with the addition of the sensitive question. They too were instructed to report the total number of questions for which they agree. By assuming that the prevalence rate of the innocuous behaviors were the same in both groups, an estimate of the sensitive attribute was made by simply subtracting the rate of positive responses between the two samples, referred to as the "difference-in-means" estimator. A more efficient form of the ICT was the double-lists version of item count (DICT; Droitcour et al., 1991; Glynn, 2013). The DICT reduced variability considerably compared to the ICT since the number of participants answering the sensitive question was doubled. In the technique, two samples of participants were used along with two sets of item lists, A and B. Each item list contained a set of innocuous or non-sensitive questions. For the first sample, participants responded to item list A, which included the addition of the sensitive question and then responded to the innocuous list of questions in item list B. The second sample received item list B with the addition of the sensitive question and then responded to the list of innocuous questions in item list A. As a result, all participants in the sample responded to the sensitive question, which doubled the number of participants responding to the sensitive question in the ICT. Because of this, the estimate of the DICT was more efficient. Recently, Petróczi et al. (2011) developed a fuzzy response model, the single sample count (SSC) technique in an attempt to simplify and provide a more economically savvy form of the ICT and DICT. In the Petróczi model, the need for an additional sample was eliminated by including innocuous questions in the item list with known probabilities of 0.50. By doing this, the model was made more efficient since all survey participants were now used to estimate the sensitive behavior instead of “wasting” a proportion of the sampled respondents to estimate the rate of the innocuous behaviors. Examples of innocuous questions used in the single sample count technique (SSC) would be “My birthday is in the first 6 months of the year”, “My house number ends with an even number,” and “The last digit of my telephone number is even”. Since the number of innocuous questions in the item list were known and have a 50-50 chance of endorsement, the estimator of the sensitive behavior could be easily derived by simply subtracting the proportion of endorsed items in the sample from the expected value of the endorsed items from the innocuous list of questions.

Even though the utilization of the item count technique (ICT) and double item count technique (DICT) had grown, the methodological research on the topic remained low. Recent methods that improve efficiency of the technique included analysis by subpopulations or domains (Tsuchiya, 2005), modification in the manner in which the sensitive item was included (Chaudhuri & Christofides, 2007), adjustments to the difference in means estimator (Glynn, 2013), correlation between non-sensitive items in the item list (Glynn, 2013), correlations between the two DICT item lists (Glynn, 2013) and the development of new nonlinear least squares and maximum likelihood estimators for multivariate analysis (Blair & Imai, 2012; Corstange, 2009; Imai, 2011).

Studies using the ICT have reported mixed results, but were generally favorable. Several studies using the technique produced estimates that appeared to reduce social desirable reporting (Holbrook & Krosnick, 2010; Rayburn, Earleywine, & Davison, 2003). Holbrook and Krosnick (2010) performed an extensive analysis of 48 studies comparing estimates between the ICT to direct questioning and found that 63% of the studies reported higher prevalence rates of the sensitive attribute from surveys utilizing the ICT compared to those using direct questioning. Although the authors concluded that the item count technique (ICT) could actually improve the validity of self-reports by reducing bias associated with social desirability pressures, their results also indicated that the technique could not be immune to forms of evasive answer bias since 27% of the studies resulted in lower or similar prevalent rates compared to the direct questioning technique. Two additional studies using the technique also produced estimates that appeared to under report prevalence rates compared to direct questioning (Biemer & Wright, 2004; Droitcour et al., 1991). In other cases negative prevalent rates have been reported for both ICT and SSC (Petróczi et al., 2011; Tsuchiya et al., 2007). In their conclusion, Droitcour et al. (1991) determined that the ICT could be problematic for sensitive behaviors with low prevalence rates such as drug use. This particularly occurs when the variability in the rate of "yes" responses taken from the sample of subjects with the additional sensitive question was inflated. Tsuchiya et al. (2007) provided the first empirical study of the effects on estimates and variation as an increasing function of the number of innocuous questions included in the ICT item list. As a general rule of thumb, most ICTs included five total questions: 4 innocuous questions and 1 sensitive question but no empirical studies have been made to verify the optimality of this number. Results of Tsuchiya et al. (2007) study indicated no changes in prevalent rates of the sensitive behavior when the number of innocuous questions in the item list increases from 2 to 5; however, variation increased with the addition of more innocuous questions. Since Tsuchiya et al.'s study included sensitive attributes (blood donation, shoplifting) with higher prevalence, the affects of list size on estimates of sensitive attributes with lower prevalence have not been formerly studied. As a result, no study has been done that would provide optimal list sizes for the item count technique (ICT), double item count technique (DICT) and single sample count technique (SSC) that maximize protection of exposure and at the same time minimizes variation.

Tsuchiya et al. (2007) considered under-reporting a potential factor in the instability of estimates using ICT. Under-reporting occurred when participants did not fully endorse the number of innocuous items for which they belong. As a result, the estimated prevalence of the sensitive attribute was under-reported. In their study, Tsuchiya et al. (2007) determined that prevalent rates from innocuous item lists questions that were asked in a direct method were higher than corresponding rates from the same item lists asked in the ICT format. For the ICT, DICT, and SSC this type of falsification distorted prevalence and thus, could be used to define a form of non-compliant responding in these types of models. The affects of such distortions were not known since non-compliant responding in ICT, DICT and SSC models has not been formerly defined or studied.

Statement of Problem

Indirect questioning techniques such as RR, ICT, DICT, and newly developed SSC were developed to elicit more truthful responses to survey questions of sensitive behaviors, resulting in less biased estimates compared to estimates elicited from direct questioning. However, estimates from these techniques were often distorted since participants were non-compliant or failed to follow the instructions of the technique. For the RR model, methods were developed to estimate non-compliance and provided an adjusted estimate of the sensitive attribute. These methods usually defined non-compliance in terms of the self-protected no response where it was assumed that non-compliant survey participants always answered in the negative regardless of the question selected. For ICT, DICT, and SSC non-compliance has not yet been formally defined or studied. Tsuchiya et al. (2007) determined that under-reporting of questions in the item list for the ICT, DICT, and SSC method resulted in underestimating the prevalence of the innocuous questions which resulted in a distorted estimate. Although studies examined the effects of under-reporting in the ICT (Blair & Imai, 2012; Corstange, 2009; Glynn, 2013), no study has officially defined non-compliance for the ICT, DICT, or SSC. Nor has there been any study that examined under what conditions the random response (RR) techniques were less sensitive to non-complaint response bias compared with non-random response (NRR) models such as the ICT, DICT, or SSC.

Several factors have affected non-compliance rates of these models. For the RR models, eliciting truthful responses was a function of the probabilities associated with selecting the sensitive question--if chosen incorrectly this parameter could encourage non-compliant responding. For the ICT, DICT, and SSC models, eliciting truthful responses was a function of the total number of questions in the item list where a longer list of questions encouraged more honest responding since the likelihood of the participant endorsing all the items was very low. Although these confidentiality parameters were necessary in eliciting truthful responses, they increased variation. Thus there was a tradeoff between bias and variability that should be considered when selecting between these techniques. As a result, the number of sampled participants was another factor that needed to be considered in measuring the effects of non-compliance.

Purpose of Study

The study examined:

1. Under what conditions was the RR models more sensitive to non-compliance compared to the item count technique (ICT), double item count technique (DICT), and single sample count (SSC)? Factors evaluated:

    a) Bias

    b) Efficiency

    c) Sample size

    d) Estimated prevalence of the sensitive attribute

    e) Confidentiality parameters

        i. probability of selecting the sensitive question (RR),

        ii. item list = the total number of innocuous questions in the item list (ICT, DICT, and SSC),

        iii. probabilities associated with the innocuous question(s) (ICT, DICT, and SSC)

        iv. correlations between the innocuous questions in the item list (ICT, DICT, and SSC)

2. Can an optimal number of innocuous questions included in the ICT, DICT, and SSC be found that maximizes compliance but minimizes additional variation? Factors evaluated

    a) Bias

    b) Efficiency

    c) Sample size

d) Estimated prevalence of the sensitive attribute

e) Item list = the total number of innocuous questions in the item list (ICT, DICT, SSC)

f) Probabilities associated with the innocuous question(s) (ICT, DICT, and SSC)

g) Correlations between the innocuous questions in the item list (ICT, DICT, and SSC).

Research Questions

Q1 Are the indirect question techniques of the ICT, DICT, and SSC models more efficient, in the presence of non-compliant reporting, as measured by their Mean Squared Error (MSE) compared to the MSE of the RR models using the unrelated question technique and forced-choice techniques?

Q2 Is there an optimal number of innocuous questions in the item list for the ICT, DICT, and SSC techniques that will reduce non-compliance and minimize additional variation?

Significance of Study

By studying the effects of distortion and the efficiency of estimates due to non-compliant responding in models that utilized indirect responses, guidelines could be developed that describe under what circumstances certain techniques would be more beneficial than others. Guidelines would include sample size calculations necessary to determine efficient estimation as well as an optimal number of innocuous questions to be included in the item list for the ICT, DICT, and SSC techniques.

Definitions

Efficiency. Used to compare statistical procedures and, in particular, refers to a measure of the optimality of an estimator by comparing the variances between estimators.

A more efficient estimator requires fewer samples than a less efficient estimator. In general, ratios of Mean Squared Errors were used to estimate efficiency.

Item list. A series of innocuous survey questions included with a sensitive question that a survey participant can either endorse or not endorse.

Non-compliance. A type of response method where participants did not follow the instructions of a survey instrument and responded in a manner that was not truthful. Other terms include “cheaters” and “self protected no response.”

Self-protected no response. A type of non-compliance that was particularly evident in RR models and resulted when participants disregarded instructions and responded negatively to a question.

CHAPTER II

REVIEW OF THE LITERATURE

Background

The chapter has been broken down into six main parts: (a) a review of benefits and limitations of random response (RR) models and non-randomr response (NRR) models compared with models using direct questioning technique (DQT); (b) an overview of the RR models, Warner's unrelated question, and forced response; (c) an overview of the NRR models, item count technique (ICT), double item count technique (DICT), and single sample count (SSC); (d) an overview of the effects of non-compliance on estimators from RR and NRR models and how these are remedied; (e) the size effects of the item question list; and (f) an overview of the generation of correlated artificial binary data.

Benefits and Limitations of Random Response and Non-random Response Models

It has been well documented in the literature that estimates of sensitive topics from surveys utilizing the direct questioning technique (DQT) where participants are asked to respond to the sensitive question directly resulted in higher non-response (Fox & Tracy 1986; Tourangeau & Yan, 2007) and higher evasive response bias (Belli et al., 2001; Fu et al., 1998; Greenberg et al., 1969; Tourangeau & Yan, 2007; Warner, 1965;). Tourangeau and Yan (2007), for instance, reviewed a series of studies comparing results from self-report illicit drug use to results from urinalyses and found that between

30%-70% of those testing positive to illicit drugs claimed they had not used drugs recently. In addition, Belli et al. (2001) found that 20% of participants of the American National Election Studies reported they had voted when they actually had not. This was determined after comparing self-report estimates to actual voting records. Fu et al. (1998) found that abortion rates were also under-reported. They compared self-report measures from the National Survey of Family Growth to data from abortion clinics and concluded that approximately 52% of total abortions were self reported.

Random response (RR) and non-random response (NRR) techniques were developed in an attempt to encourage more honest responding by providing participants with an extra level of protection. In the RR model, protection was provided by having the participant respond to one of two questions, either the sensitive question or a second question, using a randomized device. Since the interviewer was unaware of the question selected and only provided the response, the survey participant was made to feel more secure in honestly answering the sensitive question if selected. In NRR techniques (i.e., ICT, DICT, and SSC), the sensitive question was embedded in a list of innocuous questions where the participant was only asked to state the number of questions that are true. Because RR and NRR models offer these additional protections, more honest responses were expected resulting in a less bias estimate of the sensitive attribute.

The literature cites several studies where the estimate of the sensitive attribute was improved when a random response (RR) model or non-random response (NRR) model was used compared to direct response. Generally, improvements to estimators are determined if the technique produced a higher estimate compared to direct questioning when the behavior was socially unacceptable and a lower estimate when the behavior was socially acceptable (Holbrook & Krosnick, 2010; Lensvelt-Mulders., Hox, van der Heijden, & Maas, C. J., 2005). Lensvelt-Mulders, Hox, van der Heijden, and Maas. (2005) performed an extensive meta-analysis on a variety of random response techniques and sensitive topics. The authors determined that random response techniques produced better prevalence estimates of sensitive attributes compared with conventional methods such as direct questioning. In addition, response rates were generally higher when randomized techniques were used (Clark & Densharnais, 1998). Holbrook and Krosnick (2010) performed an extensive analysis of 48 studies comparing estimates between the non-random response (NRR) ICT to direct questioning and found that 63% of the studies reported higher prevalence rates of the sensitive attribute from surveys utilizing the item count technique (ICT) compared to those using direct questioning. The authors went as far to suggest that the ICT may actually improve the validity of self-reports by reducing bias associated with social desirability pressures (Holbrook & Krosnick, 2010). In another study, Rayburn et al. (2003) compared estimates using the non-random response ICT against the direct questioning approach to measure base rates for anti-gay hate crime perpetration among college students. Results indicated higher prevalence rates of "getting into a fight with someone because they are gay OR destruction of property because they were gay" among the students surveyed using the ICT compared to the rates from those who were asked to respond directly. More recently, Holbrook and Krosnick (2010) used the technique to compare estimates of voter turnout rates compared with rates from the conventional direct question method. Fielding two types of surveys, telephone and self-administered via a computer or over the internet, the authors concluded that voter turnout rates from the ICT were expectedly lower than those from direct questioning among participants from the telephone surveys. Estimates from the computer survey were closer indicating that the manner in which surveys were administered (telephone vs. computer self-administration) potentially affect self-report measures. In an attempt to validate the item count technique (ICT), Tsuchiya et al. (2007) compared estimates using the ICT and direct question technique for the behaviors shoplifting and blood donation. These behaviors were selected since they were more prevalent in the population and are opposite in terms of socially acceptable behavior. The authors hypothesized that the validity of the ICT would be verified if the estimates of the less stigmatizing behavior from both techniques were similar and a higher estimate of the stigmatizing behavior occurred when the ICT was used. Results were conclusive where the prevalence rates of blood donation were similar between the two techniques and the prevalence rates of shoplifting were approximately 10% higher among the ICT. As a result, the authors were able to conclude that the ICT was practical for research of sensitive topics.

Despite the effort to produce unbiased estimates of sensitive topics, random response (RR) and non-random response (NRR) models have several limitations compared to the direct questioning technique. For one, both models included additional sources of variation since a random device was used in the RR and an item list of innocuous questions was used in the NRR. In order to improve efficiency, each technique required larger sample sizes compared to models that use the direct questioning approach. As a result, when determining costs the experimenter must decide between accounting for less bias but additional variation and increased sample size using the RR or NRR models and a biased estimator with less variation using the direct questioning approach. Secondly, when using the RR technique, additional costs were acquired since a random device--such as spinners, coins, and cards were necessary. Reproducibility was also an issue when using the random response (RR) technique since participant's responded to questions in a random manner. For both techniques, instructions may prove difficult to follow and the participant may not respond appropriately, resulting in further distortion of the estimate. In addition, estimates from both techniques were not immune to non-compliant responses which often times distort the prevalence despite the efforts of providing the participant with extra protection. Non-compliance was especially important since the purpose of these techniques was to encourage honest responding and therefore more accurate estimation.

A review of the literature revealed several occurrences where these types of limitations resulted in mixed or problematic estimation. Shimizu and Bonham (1978), for instance, examined self-reported abortion rates from the 1973 National Survey of Family Growth (NSFG), a representative survey of women 14 to 44 years of age meant to produce national estimates for the non-institutional U.S. population. The survey used the unrelated question technique and fielded two separate samples of women to estimate abortion prevalence among married and unmarried women with children. Prevalence of the sensitive behavior from each sample differed significantly where the estimated prevalence from the first sample was 5.3 and the second was 0.6--a difference the authors reported as three times greater than the standard error of the difference. Although the randomized response technique produced a higher estimate than previously reported, the authors cautioned its use due to potential measurement errors and the additional variation associated with the randomized response technique. In another study comparing the effects of self-report between three survey techniques, self-administered, personal interviewed, and the randomized response unrelated question technique. Wiseman et al. (1975) found that the prevalence rates from the randomized response technique of four of the five sensitive questions regarding racial prejudice, political and moral issues were similar to those from the personal interview technique. In addition, the prevalence rates from both methods differed significantly compared with the less liberal estimates from the sample of participants who self -administered the survey. In order to determine the level of confidence in the randomized model, Wiseman had included a supplementary question asking participants if they were confident that the random device protected their anonymity. Since 20% of the randomized response participants felt the interviewer knew what question they selected, Wiseman concluded that distrust in the randomized technique may have been one reason the random response estimates were similar to the estimates from the personal interviews. More recently, Petróczi et al. (2011) compared prevalence rates of Mephedrone usage between the forced response technique and the NRR single sample count technique (SSC) among 318 male volunteers in north Wales and urban areas of England. Volunteers completed two surveys with each technique--in random order. In addition, approximately half of the volunteers provided hair samples in order to estimate the actual prevalence rate. Prevalence of Mephedrone usage from the forced response model was 8.81 (95% confidence interval: 2.6 and 15.00), whereas prevalence rates from the hair samples was just 4%. Having not adjusted for non-compliance rates, the authors concluded that self-protected or non-compliant responding potentially distorted their estimate. In a study that investigated intravenous drug use and receptive anal intercourse, Droitcour et al. (1991) compared the non-random response ICT to the direct question approach and found that estimates from the direct question technique were higher for the entire sample. In another study Biemer and Wright (2004) used the item count technique (ICT) to estimate the prevalence of cocaine use and found that the prevalence estimates from self-direct questioning were also higher. In their conclusion, Droitcour et al. (1991) determined that the ICT is problematic for sensitive behaviors with low rates--such as drug use. This particularly occurs when the variability of the rate of “yes” responses taken from the sample of subjects with the additional sensitive question is inflated (Droitcour et al., 1991). Despite the effort to increase honest responses, RR and NRR techniques are not immune to estimate distortions as revealed in the literature when compared with the less costly more efficient direct response technique. The question then becomes how these types of distortions affect estimation when using the RR and NRR techniques and was one method more preferable in terms of reducing bias and increasing efficiency.

Overview of Random Response Models

The random response (RR) model for proportions was first introduced in 1965 by Stanley L. Warner in his breakthrough paper, *Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias*. The idea was to provide an alternative technique, other than direct questioning, that increased cooperation and reduce what Warner defined as "evasive answer bias". In his paper, Warner (1965) defined two types of “evasive answer bias.” The first, refusal or non-response bias, occurred when the survey participant refuses to answer the sensitive question. The second, response bias or bias due to falsification, occurred when the survey participant did not answer the sensitive question honestly. Warner does not distinguish between participants who refuse to answer the question from those who choose to answer dishonestly. Instead, the method focused on reducing both types of biases collectively.

In Warner's technique, instead of asking the survey participant directly if they possess the sensitive attribute, the participant randomly selected and responded "yes" or "no" to one of two complementary questions:

1. Are you a member of population A? (i.e., π_s)
2. You are not a member of population A? (i.e., $1-\pi_s$)

where population A is the population with the sensitive attribute (π_s).

Selection of the question is made via a randomized device with a known probability (*p*), controlled by the interviewee, such as tossing a die or selecting a card. In Warner's model, a spinner was used which selected the first question with probability *p* and the second with probability 1-*p*. Since the survey participant controlled the device, the interviewer or researcher was unaware of the question selected, only the response given. Warner argued that the method theoretically increased honest reporting since the participant was less likely to respond to the sensitive question. The likelihood of answering the sensitive question depended on *p*. For instance, in the direct questioning technique, the probability of answering the sensitive question is 1 and therefore "yes" implies the participant possessed the sensitive attribute; whereas in Warner's RR technique, since the probability of answering the sensitive question is less than 1, a participant who reported "yes" may or may not possess the sensitive attribute.

The model as defined by Warner (1965):

Let

π = the true probability of A in the population,
p = the probability the random device selects A, and

$$X_i = \begin{cases} 1 & \textit{if the ith sample subject responds yes} \\ 0 & \textit{if the ith sample subject responds no} \end{cases}$$

Then

$$P\{X_i = 1\} = \pi\ p + (1 - \pi)(1 - p),$$
$$P\{X_i = 0\} = (1 - \pi)\, p + \pi(1 - p),$$

arranging the indexing of the sample so that the first n_1 report “yes” and the second (n - n_1) report "no", the likelihood of the sample was:

$$L = [\pi p + (1-\pi)(1-p)]^{n_1}[(1-\pi)p + \pi(1-p)]^{n-n_1} \tag{2.1}$$

and the log of the likelihood was:

$$log(L) = n_1 log[\pi p + (1-\pi)(1-p)] + (n-n_1) log[(1-\pi)p + \pi(1-p)].$$

Taking the first derivative and setting this to zero obtained:

$$\frac{dlog(L)}{d\pi} = \frac{n_1}{\pi p + (1-\pi)(1-p)}\{(2p-1)\} - \frac{n-n_1}{(1-\pi)p + \pi(1-p)}\{(2p-1)\}$$

and necessary conditions on π for a maximum were

$$\frac{(n-n_1)(2p-1)}{(1-\pi)p + \pi(1-p)} = \frac{n_1(2p-1)}{\pi p + (1-\pi)(1-p)} \tag{2.2}$$

which reduced to

$$\frac{2p-1}{\pi p + (1-\pi)(1-p)} \frac{(1-\pi)p + \pi(1-p)}{2p-1} = \frac{n-n_1}{n_1}$$

$$\frac{(1-\pi)p+\pi(1-p)}{\pi p+(1-\pi)(1-p)}=\frac{n}{n_1}-1$$

$$\frac{(1-\pi)p+\pi(1-p)}{\pi p+(1-\pi)(1-p)}+\frac{\pi p+(1-\pi)(1-p)}{\pi p+(1-\pi)(1-p)}=\frac{n}{n_1}$$

$$\frac{1}{\pi p+(1-\pi)(1-p)}=\frac{n}{n_1}$$

$$\pi p+(1-\pi)(1-p)=\frac{n_1}{n}. \tag{2.3}$$

Then, supposing $p \neq \frac{1}{2}$, the maximum likelihood estimate of π was

$$\hat{\pi}=\frac{p-1}{2p-1}+\frac{n_1}{(2p-1)n}. \tag{2.4}$$

and since $n_1=\sum_{i=1}^{n}(X_i)$ the expected value of $\hat{\pi}$ was:

$$E[\hat{\pi}]=\frac{1}{2p-1}\left[p-1+\frac{1}{n}\sum E(X_i)\right]$$

$$=\frac{1}{2p-1}\left[p-1+\pi p+(1-\pi)(1-p)\right]$$

$$=\pi, \tag{2.5}$$

and the variance of $\hat{\pi}$ was

$$Var(\hat{\pi})=\frac{nVarX_i}{(2p-1)^2n^2}$$

$$=\frac{[\pi p+(1-\pi)(1-p)][(1-\pi)p+\pi(1-p)]}{(2p-1)^2n}$$

$$= \frac{\frac{1}{4} + (2p^2 - 2p + \frac{1}{2}(-2\pi^2 + 2\pi - \frac{1}{2}}{(2p-1)^2 n}$$

$$= \frac{1}{n}\left[\frac{1}{16(p-\frac{1}{2})^2} - (\pi - \frac{1}{2})^2 \right]. \tag{2.6}$$

Expression (2.6) also set out the separate dependence of the variance of $\hat{\pi}$ upon the choice of p. Identifying

$$\frac{\frac{1}{4} - (\pi - \frac{1}{2})^2}{n} = \frac{\pi(1-\pi)}{n}$$

as the variance due to sampling and "adding" and "subtracting" $\frac{1}{4}$ from both equations in expression (2.6), the variance of $\hat{\pi}$ can be rewritten as

$$Var(\hat{\pi}) = \frac{\frac{1}{4} - (\pi - \frac{1}{2})^2}{n} + \frac{\frac{1}{16(p-\frac{1}{2})^2} - \frac{1}{4}}{n},$$

$$= \frac{\pi(1-\pi)}{n} + \frac{\frac{1}{16(p-\frac{1}{2})^2} - \frac{1}{4}}{n}. \tag{2.7}$$

From this, it is clear that the variance of $\hat{\pi}$ can be expressed as the sum of the variation due to sampling and the variation due to the random device.

As is evident in expression (2.5), $\hat{\pi}$ is an unbiased estimator of π and since $\hat{\pi}$ is a maximum likelihood estimator, when *n* is large, can be assumed normally distributed. As a result, confidence intervals can be formed as usual.

The selection of the parameter p was crucial to the model. Since the variability of π can be decomposed into the variability due to the sample and the additional variability due to the selection of p (2.7), it is not hard to see that as p moves from the endpoint (i.e., either 0 or 1) toward 0.50, the additional variation due to the parameter approaches maximum. Warner (1965) pointed out that this additional variation is necessary since in theory its purpose was meant to increase participant cooperation which in turn reduced evasive answering bias. However, with the additional variation, the precision of the estimate as compared to the estimate from direct questioning declined unless the sample size using the RR technique was increased. For instance, as demonstrated by Warner (1965), suppose $\pi = 1/2$ and $p = 3/4$, then the variance shown in (2.6) is $1/n$. For an estimate with a standard deviation of 0.05, this implies a sample size of 400; whereas in the direct questioning technique (equivalent to $p = 1$), would take a sample of approximately 100 subjects. Thus, in Warner's model, although p was meant to increase cooperation and, therefore, reduce evasive answer bias, it came at a cost since p also increased model variability. Therefore, when using the RR technique, an experimenter must decide if the reduction in bias is worth the additional cost of variation.

In his original paper, Warner (1965) demonstrated that in certain situations the RR technique was superior to the technique of direct questioning. Under the assumption of truthful reporting when the RR technique was used and “less truthful” reporting when the direct questioning technique was used, Warner compared the mean squared errors of both techniques in a simulation study. For estimates of the direct questioning technique, Warner allowed for a combination of different levels of truthfulness from participants with the sensitive attribute (proportion of truthful responses ranged from 50% to 100%)

as well as those without the sensitive attribute (proportion of truthful responses ranged from 50% to 100%). For estimates of the RR technique, only truthful responses were assumed. Simulations for both techniques were calculated for selected values of *p* (0.60, 0.70, 0.80, 0.90), fixed sample size (*n* = 1,000) and a sensitive prevalence of π = 0.50 and of π = 0.60. In his demonstration, Warner supposed a group of participants agreed to be surveyed.

For estimates of the direct questioning technique, he defined:

T_α = Probability that the members of group A (sensitive population) tell the truth

T_β = Probability that the members of group B (complement population) tell the truth

Recalling:

$$X_i = \begin{cases} 1 & \textit{if the ith sample subject responds yes} \\ 0 & \textit{if the ith sample subject responds no} \end{cases}$$

then the conventional estimate of the true population proportion $\hat{\pi}$ was

$$\hat{\pi} = \frac{\sum_{i=1}^{n} X_i}{n}. \quad (2.8)$$

and the expected value, bias and variance given by (Warner, 1965)

$$E[\hat{\pi}] = \pi T_\alpha + \lfloor (1-\pi)(1-T_\beta) \rfloor \quad (2.9)$$

$$Bias[\hat{\pi}] = E(\hat{\pi} - \pi)$$

$$= \pi \lfloor T_\alpha + T_\beta \rfloor + \lfloor 1 - T_\beta \rfloor, \textit{ and} \quad (2.10)$$

$$Var\hat{\pi} = \frac{\lfloor \pi T_\alpha + (1-\pi)(1-T_\beta) \rfloor \lfloor 1 - \pi T_\alpha - (1-\pi)(1-T_\beta) \rfloor}{n} \tag{2.11}$$

Under the assumption that RR participants told the truth, Warner's simulations demonstrated that the RR technique was more efficient if it is expected that between 30% to 50% of direct questioning participants in either population did not tell the truth.

Greenberg et al. (1969) demonstrated that despite Warner's effort to reduce evasive response bias, in cases of more stigmatizing behaviors (i.e., $\pi = 0.05$ or $\pi = 0.10$) where the probability of the random device low ($p = 0.05$, $p = 0.20$), unless false responding was high (i.e., between 25% and 50% of participants not responding truthfully), Warner's technique proved less efficient compared to the direct question technique. Greenberg et al. (1969) selected prevalence levels that more accurately reflected socially stigmatizing behaviors (i.e., induced abortion, drug addiction) compared to the attribute rates selected by Warner ($\pi = 0.50$ and of $\pi = 0.60$) - which are more descriptive of less stigmatizing activities such as voting behaviors. As pointed out by Greenberg et al. (1969), if membership into Group A (i.e., π) was small (i.e., $\pi = 0.05$) and the selection of the sensitive question was also smaller (i.e., $p = 0.05$), holding the sample size fixed at 1,000 (i.e., no non-response bias) and assuming participants answer honestly, Warner's technique was only 1/2 as efficient as the technique of direct questioning and only 1/10 as efficient if $p = 0.20$. If, on the other hand, members of group A responded truthfully only 90% of the time, Warner's technique was 2/3rds as efficient as direct questioning when $p = 0.05$ and 1/7th as efficient when $p = 0.20$. As evasive responding increased, Warner's technique became superior. For instance, if 25% of members of group A do not respond honestly, Warner's technique was two times as efficient compared with the direct questioning technique when $p = 0.05$ and 1/2 as efficient when $p = 0.20$. At 50% falsification, Warner's technique was six times as efficient for $p =$

0.05 and 4/3rds as efficient for $p = 0.20$. Thus, for socially stigmatizing behaviors with smaller prevalence rates, Greenberg et al. (1969) demonstrated that the Warner technique only achieved superiority when a large proportion of members were expected to respond dishonestly. Because of this, the authors concluded that a more efficient random response (RR) technique was necessary since, in the presence of both non-response and/or large response falsification, using the direct questioning technique to estimate the sensitive attribute was inappropriate. As a result, the authors presented the theoretical framework of the RR model developed by Walt A. Simmons (Horvitz et al., 1967) which was known as the unrelated question technique (UQT).

Recall, that both questions posed in the Warner model:

1. Are you a member of population A (i.e., π_s)?
2. You are not a member of population A (i.e., $1 - \pi_s$)?

referred to the sensitive population, π_s. Because of this, participants may not cooperate as fully as Warner believed. For instance, if a member of population A were to select question 2, they may feel less secure in responding “no” since this would imply they were a member of the sensitive population. To address this issue, Simmons suggested posing the sensitive question with an unrelated or innocuous question such as, “were you born in the first half of the year?,” “are you left handed?,” or “were you born in the state of Colorado?” Thus, the two questions posed in the UQT were:

1. You are a member of population A, (i.e., π_s)
2. You are a member of population Y, (i.e., π_{ns})

As was the case in Warner’s model, the participant responded to a question they selected using a random device (such as Warner’s spinner). However, the difference, if the second question was selected, a participant who was a member of population A, would feel less embarrassed responding “no” since population Y was independent or unrelated to population A.

Another difference between the unrelated question technique (UQT) and Warner’s technique was the additional parameter, π_{ns} which also must be estimated. This could be done in two ways: (a) where π_{ns} is unknown (i.e., were you born in the state of Colorado and (b) where π_{ns} was known or approximately known (i.e., were you born during the first half of the year?). The former case is discussed first.

In the case of estimating unknown π_{ns} two independent non-overlapping samples were necessary. According to Greenberg et al. (1969), sample sizes did not need to be equal and could actually be made unequal to produce more efficient estimation. Participants from each sample were provided with the same two questions, the sensitive question and an innocuous question. Two random devices, one for each sample, were used to select between the questions where the probability of selecting the sensitive question in the first sample, p_1, is different from the probability of selecting the sensitive question in the second sample, p_2.

Let p_1 be the probability that statement A was selected by the random device in the first sample and let p_2 be the probability statement A was selected by the random device in the second sample where $p_1 \neq p_2$. Similarily, let $(1-p_1)$ be the probability that the random device selected statement Y in the first sample, that $(1-p_2)$ be the probability of selecting statement Y in the second sample.

Under the assumption that respondents report with 100% truthfulness, let

$$\lambda_1 = \textit{the probability that a 'Yes' answer will be reported in the first sample}$$

$$= p_1\pi_s + (1 - p_1)\pi_{ns} \quad (2.12)$$

Similarly, let

$$\lambda_2 = \textit{the probability that a 'Yes' answer will be reported in the second sample}$$

$$= p_2\pi_s + (1 - p_2)\pi_{ns} \quad (2.13)$$

One can construct the liklihood function as was done in Warner's model above. Likewise the identical value was also obtained by solving (2.12) and (2.13) for $(\pi_s)_u$ where *u* distinguished the estimate of the UQT

$$(\pi_s)_u = \frac{\lambda_1(1 - p_2) - \lambda_2(1 - p_1)}{p_1 - p_2}, \quad (2.14)$$

where $p_1(1 - p_2) - p_2(1 - p_1) \neq 0$ since $p_1 \neq p_2$

A similar expression was dreived for π_{ns} as

$$(\pi_{ns})_u = \frac{p_2\lambda_1 - p_1\lambda_2}{p_1 - p_2}. \quad (2.15)$$

Now, define

$$\hat{\lambda}_1 = \frac{n_1'}{n_1} \; and,$$

$$\hat{\lambda}_2 = \frac{n_2'}{n_1}$$

where n_1' was the total number of "yes" responses from sample 1 and likewise, n_2' was the total number of "yes" responses from sample 2.

Thus, the sample estimate becomes:

$$(\hat{\pi}_s)_u = \frac{\hat{\lambda}_1(1-p_2)-\hat{\lambda}_2(1-p_1)}{p_1-p_2} \tag{2.16}$$

The observed proportions, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, are distributed as binomial random variables with parameters (n_1, λ_1) and (n_2, λ_2), respectively. It therefore follows that the expression in (2.16) was unbiased and its variance was given by:

$$Var(\hat{\pi}_s)_u = \frac{1}{(p_1-p_2)^2}\left\{\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1}+\frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right\}. \tag{2.17}$$

Estimates of the variation were made by substituting $\hat{\lambda}_1$ and $\hat{\lambda}_2$ in expression (2.17).

Greenberg et al. (1969) outlined a method in which the selection of the model parameters p_1, p_2, π_{ns}, n_1, n_2 can be made in such a manner as to attain a variance as close to the binomial distribution as possible. Note that the denominator of (2.16) was very small when p_1 is selected to be close to p_2, which would produce an estimate of π_s greater than unity. In order to ensure this does not happen, Greenberg et al. (1969) suggested an optimal choice of the randomized probabilities by selecting a p_1 and p_2 as far from each other as possible. This is best accomplished when $p_1 + p_2 = 1$ where such a choice has the additional benefit of affecting each sample in an identical but complementary manner by the randomized device. At the same time, the individual selection of each probability parameter should be made as far as possible from 0.50 which minimizes the additional variation due to both randomization devices. Greenberg et al. (1969) recommended selecting p_1 at 0.80 or 0.20, $\pm$ 0.10, which, in turn produces a similar but complementary

p_2. Since the selection of both parameters would be closer to the parameter endpoints (0, 1), the additional variation was minimized. Of course, first and foremost, both parameters needed to be selected in such a way as to encourage honest responding. In addition, selection of prevalence of the innocuous behavior, π_{ns}, should also be considered and was based on the expected prevalence of the sensitive behavior. Greenberg et al. (1969) recommended selecting π_{ns} on the same side of 0.50 as the expected prevalence of the sensitive behavior, π_s but as far from 0.50 as possible--again in an effort to minimize variation. However, when selecting π_{ns}, the researcher must keep in mind compromising the cooperation of the survey participant. If the expected prevalence of the sensitive behavior were small, say 0.05 and the innocuous behavior was also small, say 0.10, the likelihood of a "yes" response would therefore be small. As a result, the participant with the sensitive attribute could feel less secure in responding "yes"--since a "yes" response appears suspect. In determining optimal sample sizes, Greenberg provided a formula based on the selection of the model parameters p_1, p_2, and π_{ns}.

Although selecting the parameters in this fashion was meant to minimize variation, Greenberg noted that compromising the cooperation of the participant must be the researcher's primary concern. If for instance, a selection of p_1 closer to 0.50 was sufficient to guarantee cooperation, then selecting a p_1 in accordance to recommendations above could compromise cooperation and encourage evasive answering. As a result, model parameters should be selected in a manner that maximized both confidentiality and statistical efficiency.

In order to compare the loss or gain of introducing the unrelated characteristic into the RR model, Greenberg et al. (1969) compared the MSE from the unrelated question technique (UQT) to the MSE from the Warner technique under the assumption of differing levels of non-compliance or not completely truthful reporting. In their simulation study, Greenberg et al. (1969) assumed non-compliance among the sensitive population (population A) and defined non-compliance occurring exclusively among members of population A who respond "no" when in actuality should have responded "yes." Members of population Y were assumed to tell the truth. The simulation parameters were defined as follows:

Let:

$$Ta_u = Pr[\text{Member of Group A tells the truth in UQT in either sample 1 or sample 2}]$$
$$Ta_w = Pr[\text{Member of Group A tells the truth in Warner's technique}]$$

where $0 \le Ta_u \le 1$ and $0 \le Ta_w \le 1$, and where $Ta_u \le Ta_w$

From the statements above, note that Greenberg et al. (1969) assumed respondents of the Warner technique were either *less* likely or *equally* likely to respond truthfully compared with respondents of the UQT. This assumption was justified by the authors since in theory, the UQT contained the innocuous second question which, unlike the Warner technique, did not refer to the sensitive attribute. As a result, it was expected that a higher proportion of UQT respondents who are not members of population Y would respond truthfully (i.e., "no") to the second innocuous question compared with members of population A (i.e., sensitive population) who selected the second question in the Warner's technique--since a response of "no" to this question implied membership into population A.

Under these assumptions, equation (2.12) and (2.13) became:

$$\lambda_1^{'} = p_1(\pi_s Ta_u) + (1 - p_1)\pi_{ns} \tag{2.18}$$

Similarly, let

$$\lambda_2^{'} = p_2(\pi_s Ta_u) + (1 - p_2)\pi_{ns} \tag{2.19}$$

and the expected value of the estimate in (2.16) was

$$E(\hat{\pi}_s^{'})_u = \frac{\lambda_1^{'}(1 - p_2) - \lambda_2^{'}(1 - p_1)}{p_1 - p_2}. \tag{2.20}$$

The bias in this estimate was

$$\begin{aligned} Bias(\hat{\pi}_s^{'})_u &= E\{\hat{\pi}_s^{'} - \pi_s\}_u \\ &= \frac{(\lambda_1^{'} - \lambda_1)(1 - p_2) - (\lambda_2^{'} - \lambda_2)(1 - p_1)}{p_1 - p_2} \\ &= \pi_s(Ta_u - 1). \end{aligned} \tag{2.21}$$

In the case of Warner's estimate, the authors assumed non-compliance only occurred for members of population A. In this case, since both questions referred to the sensitive attribute, false reporting was assumed in the negative when members were asked the sensitive question (i.e., Are you a member of population A?) and the affirmative when members were asked the complementary question (i.e., You are not a member of population A?). Respondents who were not members of population A were assumed to respond truthfully. Thus, the probability of responding in the affirmative for the Warner method was:

$$\begin{aligned} \lambda^{'} &= \pi_s pTa_w - \pi_s(1 - p)(1 - Ta_w) + (1 - \pi_s)(1 - p) \\ &= \pi_s\left[pTa_w - (1 - p)(1 - Ta_w)\right] + (1 - \pi_s)(1 - p) \end{aligned} \tag{2.22}$$

Note, the value of (2.22) would be used as $\frac{n_1}{n}$ in (2.4) to obtain $(\hat{\pi}_s)_w$. Therefore the bias was measured by:

$$Bias(\hat{\pi}_s^{'})_w = E\{\hat{\pi}_s^{'} - \pi_s\}_w$$

$$= \frac{1}{2p-1}\left\{(p-1) + \pi_s PTa_w + \pi_s(1-p)(1-Ta_w) + (1-\pi_s)(1-p) - \pi_s(2p-1)\right\}$$

$$= \frac{1}{2p-1}\left\{(p-1)\pi_s Ta_w + \pi_s pTa_w - \pi_s(2p-1)\right\}$$

$$= \pi_s(Ta_w - 1). \tag{2.23}$$

Note that for equations (2.21) and (2.23) in the case of truthful reporting (i.e., Ta_u and Ta_w are both 1), the bias was 0; whereas in the case of less than truthful reporting (i.e., Ta_u and Ta_w are both < 1) the bias was negative. This implied that when non-compliance was defined in these terms, both techniques underestimated π_s.

Using these equations, MSEs were simulated and efficiency compared between the two techniques by the authors for decreasing values of Ta_u and Ta_w (1.00, 0.90, 0.80, 0.70, 0.60, 0.50), fixed sample size of 1,000 (where n_w = 1,000, n_{1uqt} and n_{2uqt} were selected optimally), π_s = 0.20, and $p_w = p_{1uqt}$ = 1-p_{2uqt} = 0.20. Results indicated that when the amount of truthful reporting was equal or more likely to occur in the unrelated question technique (UQT) compared with Warner's technique, UQT was superior. Since MSE = $(\text{bias})^2$ + variance, as demonstrated by the authors, as the amount of untruthful reporting increased, so too did the contribution by the bias to the MSE. The authors therefore concluded that if the UQT improved even slightly more truthful reporting compared to Warner's technique, the UQT would result in a more efficient MSE.

Suppose instead, the experimenter posed a second question with a known or approximately known population proportion, π_{ns}. Since π_{ns} would be known, only one sample would be sufficient to estimate π_s. This in turn would reduce costs as well as the additional variability necessary in estimating π_{ns}. Examples of such questions would be, “were you born in the second half of the year?” or “are you left handed?” Prevalence estimates of these questions would be obtained from other data sources, such as the Census. Since only one parameter were to be estimated, using (2.12) define:

$$\lambda_1 = \textit{the probability that a 'Yes' answer will be reported in the sample}$$

$$= p_1\pi_s + (1-p_1)\pi_{ns} \tag{2.24}$$

and the maximum likelihood estimator and sample variance as defined by Greenberg et al. (1969) were:

$$(\pi_s \mid \pi_{ns})_u = \frac{\lambda_1 - \pi_{ns}(1-p_1)}{p_1} \tag{2.25}$$

$$Var(\hat{\pi}_s \mid \pi_{ns})_u = \frac{\lambda_1(1-\lambda_1)}{np^2}. \tag{2.26}$$

Note, estimates would be obtained by substituting $\hat{\lambda}_1$ in each equation.

Greenberg et al. (1969) demonstrated that when the proportion of π_{ns} was known, the efficiency of the UQT estimate was superior to the estimate from both Warner's and the UQT when π_{ns} was not known. However, this assumed that known π_{ns} was without error, which in many cases was not accurate. For instance, an estimate using Census data of the population born during the first half of the year could include the entire U.S.

population. If the sample were drawn from residents of a specific state, say Colorado, this proportion may differ. As a result, the deviation from the true π_{ns} introduced additional bias into the estimator:

Defining π^*_{ns} as unrelated question technique (UQT) estimate of π_{ns}, where π^*_{ns} was obtained from another data source such as the Census. Assume further that $\pi_{ns} = \pi^*_{ns} + C$. Then (2.25) becomes:

$$(\hat{\pi}_s \mid \pi^*_{ns})_u = \frac{\lambda_1 - \pi^*_{ns}(1-p_1)}{p_1}. \qquad (2.27)$$

Continuing as before, with Ta_u being the probability that respondents in the sensitive population responded truthfully,

$$\begin{aligned} &Bias(\hat{\pi}_s \mid \pi^*_{ns})_u = E\{(\hat{\pi}_s \mid \pi^*_{ns}) - \pi_s\} \\ &= \frac{1}{p_1}\{(p_1\pi_s Ta_u + \pi_{ns}(1-p_1) - (\pi_{ns} - C)(1-p_1) - p_1\pi_s\} \\ &= \pi_s(Ta_u - 1) + \frac{C(1-p_1)}{p_1} \qquad (2.28) \\ &= \beta_S + \beta_{ns}. \end{aligned}$$

where

β_s is the bias due to untruthful reporting of the sensitive characteristic and

β_{ns} is the bias due to the erroneous estimation of π_{ns} from sources external to the survey.

Greenberg et al. (1969) presented findings that demonstrated when π_{ns} is approximately known, underestimating π_{ns} resulted in a more efficient estimate compared to the Warner estimate.

Theoretically estimating π_{ns} in the unrelated question technique (UQT) by either use of two samples or using outside data was never necessary if the unrelated question were incorporated into the randomized device. This technique was first suggested by Richard Morton of the University of Sheffield (Greenberg et al., 1969) and became known as the Forced Choice Technique (FCRT). The technique uses three statements, each selected with separate probabilities that add to unity: $p_1 + p_2 + p_3 = 1$. The first statement was the sensitive question, and was selected with probability p_1, where the participant would be instructed to answer the question honestly. The second and third statements were non-sensitive statements, selected with probability p_2 and p_3, where participants were forced to respond “no” or “yes,” respectively. For instance the sum of two dice could be used to determine the appropriate response where, if a sum between 5 and 10 were observed, survey participants would be instructed to respond to the sensitive question with probability 3/4ths. If instead the sum were between 2 and 4 the participant would be instructed to respond “yes” with probability 1/6th, and if the sum were greater than or equal to 11, the participant would be instructed to respond “no” with probability 1/12th. From this, it follows that:

$$\begin{aligned}\pi_{ns} &= \frac{p_2}{(p_2 + p_3)} \\ &= \frac{p_2}{(1 - p_1)} \qquad (2.29)\end{aligned}$$

and the maximum likelihood estimator of the π_s given π_{ns} was

$$\hat{\pi}_s \mid \pi_{ns} = \frac{\hat{\lambda}_1 - (1-p_1)\pi_{ns}}{p_1}, \\ = \frac{\hat{\lambda}_1 - p_2}{p_1} \tag{2.30}$$

and the maximum likelihood estimator of the sample variance was:

$$var(\hat{\pi}_s \mid \pi_{ns}) = \frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{np_1^2}. \tag{2.31}$$

When π_{ns} was incorporated into the randomized device, it would be truly known. As a result, no additional variation or bias would be present in the estimate. Because of this, the FCT, which was a derivative of the UQT, was the most efficient of the RR techniques (Fox & Tracy, 1986; Greenberg et al., 1969).

The unrelated question technique (UQT) and forced choice technique (FCT) therefore improved on the RR model first developed by Warner. Greenberg et al. (1969) demonstrated that these techniques potentially reduced evasive response bias that is present in both the Warner's technique and direct questioning, since the additional questions used in the UQT and FRT was unrelated to the sensitive attribute. This in turn encouraged and therefore theoretically increased more honest reporting which decreased bias. In addition, the variation of the UQT and FRT proved more efficient than Warner's technique even in the presence of less than truthful reporting. As a result, the UQT and FCT were two of the more popular techniques used to estimate sensitive attributes.

Although random response (RR) models were widely used, they had limiting factors. For one, the technique can produce prevalent estimates outside the (0, 1) range. This was particularly true of the unrelated question technique (UQT) when model parameters are not defined properly (Greenberg et al., 1969). Secondly, RR models must account for the additional variation due to the randomized device and therefore, in order to produce as efficient estimates as direct questioning, sample sizes were generally larger. Additional costs were also necessary since a randomized device was used--such as dice, spinners or selecting cards. random response (RR) models contained instructions that were more difficult to follow compared to surveys that ask a question directly. Since subjects found the instructions difficult to follow, misreporting would result. Reproducibility was also an issue with RR models since responses were always randomized. Lastly and more importantly despite their best efforts, evidence of non-compliance continued to plague the estimates causing bias and inefficiencies.

Overview of Non-randomized Response Models

A second class of indirect response techniques was termed non-randomized response (NRR) models. These models utilized an even more evasive response method where survey participants never directly responded to the sensitive question. The item count technique (ICT) was probably the most widely used of the NRR models. Also referred to as the list technique and more commonly, the unmatched count technique (UCT), it was empirically demonstrated in a study by Dalton et al. (1994) but was first introduced by Miller (1984). The technique was closely related to the random response (RR) UQT where participants were randomized into two groups or samples. In the ICT however, participants from the first sample, referred to as the intervention group, were given an item list of innocuous questions including the additional sensitive question and the second sample, referred to as the control group, were given the same list of innocuous questions excluding the sensitive question. Participants were asked to read over each of the questions in the item list and respond by writing down the count of the total number of questions in the item list for which they belong. For instance, if three questions were asked in the item list, "I am left handed," "I was born in the month of November," and "I have utilized Performance Enhancing Drugs (PEDs)," and the participant belonged to two of these groups, would respond with "2." As a result, the sensitive behavior would never be tied to the participant's response since it was not known for which of the two questions the participant was a member. Thus, the expectation was that survey participants would answer more honestly.

Using this technique, what was referred to as the "difference-in-means" estimate was used to estimate the prevalence of the sensitive attribute (Glynn, 2013). By subtracting the mean number of item responses between the two samples, an estimate of the proportion with the sensitive attribute was made since the average number of item responses from the sample that were not asked the sensitive question would differ by the estimated proportion of those with the sensitive attribute from the sample that was asked the sensitive question.

For ICT, the difference in means estimator was an unbiased estimate of π_s when the following two assumptions were met (Blair & Imai, 2012; Glynn, 2013):

Letting:

$$Z_{ij}(t) = \begin{cases} 1 & \textit{if the ith sample subject responds yes to item } j \\ 0 & \textit{if the ith sample subject responds no to item } j \end{cases}$$

for $j = 1...J$ non-sensitive questions in the item list under treatment group status $t = 1$ (intervention group receiving the sensitive question in the item list) and $t = 0$ (control group receiving only the non-sensitive questions in the list)

Assumption 1: (No design effect). For each $i = 1.....N$ assume

$$\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1). \tag{2.32}$$

Assumption 2: (No liars). For each $i = 1....N$ assume

$$Z_{i,J+1}(1) = Z^{*}_{i,j+1}. \tag{2.33}$$

where $Z^{*}_{i,J+1}$ represents a truthful answer to the sensitive item.

Assumption 1 (2.32) assumed that the presence of the sensitive question in the list set of innocuous questions did not change the response patterns to the non-sensitive items between the intervention and control samples. It made no assumption about whether or not responses to the set of innocuous questions were truthful, only that the responses were similar between the two groups of participants (Blair & Imai, 2012). Assumption 2 (2.33) on the other hand assumed that participants responded truthfully to the sensitive question.

Given Assumptions 1 and 2, and defining $j = 1....J$ non-sensitive questions for the $i = 1.....n_t$ subjects in the $t = 0,1$ group, where 0 indicates the control group *not receiving* the sensitive question and 1 indicates the intervention group *receiving* the sensitive question, then

$$X_{1,J} = \sum_{i=1}^{n_1}\sum_{j=1}^{J} Z_{ij}(1). \tag{2.34}$$

was the sum of the $j = 1....J$ non-sensitive questions for the $i = 1.....n_1$ subjects in intervention group,

$$X_{1,J+1} = \sum_{i=1}^{n_1}\sum_{j=j+1}^{J+1} Z_{i,j+1}(1). \tag{2.35}$$

was the sum of the "yes" responses to the sensitive question for the $i = 1.....n_1$ subjects in the intervention group and,

$$X_{0,j} = \sum_{i=1}^{n_0}\sum_{j=1}^{J} Z_{ij}(0). \tag{2.36}$$

was the sum of the $j = 1....J$ non-sensitive questions for the $i = 1.....n_0$ subjects in the control group.

An unbiased estimator of π_s was therefore defined as:

$$\hat{\pi}_s = \frac{X_{1,J} + X_{1,J+1}}{n_1} - \frac{X_{0,J}}{n_0}$$

$$= \hat{X}_1 - \hat{X}_0 \tag{2.37}$$

The variance of $\hat{\pi}_s$ was

$$var(\hat{\pi}_s) = \frac{1}{n_{s1}}[var(X_{t=1,J} + var(X_{t=1,J+1})$$

$$+ 2Cov(X_{t=1,J}, X_{t=1,J+1})] + \frac{1}{n_{t=0}} var(X_{2t=0,J}) \quad (2.38)$$

Note that the variance of the estimator (2.38) was quite high, especially when the correlation between the sensitive question and the item list of non-sensitive items was positive. The correlation between the sensitive item and the item list of non-sensitive questions was the measure of the design effect (Blair & Imai, 2012; Glynn, 2013). If the covariance or correlation were zero no design effect was apparent, and responses to the innocuous item list were assumed similar between the two samples. If the correlation or covariance were positive or negative then a design effect was assumed and the responses to the non-sensitive items differed between the two samples. Although Blair and Imai (2012) derived a rudimentary statistical procedure to detect the presence of a design effect, no other studies have been done that either adjusted or explored the effects of the design effects in the presence of non-compliance (Blair & Imai, 2012). Another reason for high variability of the ICT was due to the fact that only one sample of subjects were asked the sensitive question (Glynn, 2013). Another more efficient version of the ICT was the double-lists version of item count (DICT; Droicteur et al., 1991; Glynn, 2013). The DICT reduced variability considerably by doubling the number of participants answering the sensitive question. For this technique, the study defined s_1 as the set of participants from the first sample and s_2 as the set of participants from the second sample. The method used two separate item count estimates taken from each sample X_{s1}^A and X_{s2}^A

and X_{s1}^{B} and X_{s2}^{B} where X_{s1}^{A} contained a series of innocuous questions along with the sensitive question and X_{s2}^{A} contained the same series of innocuous questions excluding the sensitive question. Likewise for X_{s2}^{B} and X_{s1}^{B}, which contained a different item list of innocuous questions. Sample 1 then responded to item list X_{s1}^{A} and X_{s1}^{B} and Sample 2 responded to item list X_{s2}^{A} and X_{s2}^{B}. Under Assumptions 1 (2.32) and 2 (2.33), an unbiased estimator of π_s was made:

Letting:

$$Z_{Aij}(t) = \begin{cases} 1 & \textit{if the ith sample subject responds yes to item j on list A} \\ 0 & \textit{if the ith sample subject responds no to item j on list A} \end{cases}$$

and,

$$Z_{Bij}(t) = \begin{cases} 1 & \textit{if the ith sample subject responds yes to item j on list B} \\ 0 & \textit{if the ith sample subject responds no to item j on list B} \end{cases}$$

and defining:

$$X_{sl,J}^{K} = \sum_{i=1}^{n_{si}} \sum_{j=1}^{J} Z_{ij}(1). \tag{2.39}$$

as the sum of the $j = 1....J$ non-sensitive questions for the $i = 1.....n_{sl}$ subjects from the sample $l = 1,2$ receiving the sensitive question from list $K = A, B$,

$$X_{sl,J+1}^{K} = \sum_{i=1}^{n_{si}} \sum_{j=j+1}^{J+1} Z_{i,j+1}(1). \tag{2.40}$$

as the sum of the "yes" responses to the sensitive questions for the $i = 1.....n_{sl}$ subjects from the sample $l = 1,2$ receiving the sensitive question from list $K = A, B$ and,

$$X_{sl,J}^{K} = \sum_{i=1}^{n_{si}} \sum_{j=1}^{J} Z_{ij}(0). \tag{2.41}$$

as the sum of the $j = 1....J$ non-sensitive questions for the $i = 1.....n_{sl}$ subjects from the sample $l = 1,2$ not receiving the sensitive question from list $K = A, B$.

than the unbiased estimators of π_s^A and π_s^B were:

$$\hat{\pi}_s^A = \frac{X_{s1,j}^A + X_{s1,j+1}^A}{n_{s1}} - \frac{X_{s2,j}^A}{n_{s2}}$$

$$= \hat{X}_{s1}^A - \hat{X}_{s2}^A \tag{2.42}$$

$$\hat{\pi}_s^B = \frac{X_{s2,j}^B + X_{s2,j+1}^B}{n_{s2}} - \frac{X_{s1,j}^B}{n_{s1}}$$

$$= \hat{X}_{s2}^B - \hat{X}_{s1}^B, \tag{2.43}$$

where n_{s1} was the number of participants in the first sample receiving the sensitive question in list A (likewise receiving the list of innocuous questions in list B) and n_{s2} was the number of participants in the second sample receiving the sensitive question in list B (likewise receiving the list of innocuous questions in list A).

Averaging these equations, the unbiased estimator for π_s was defined as:

$$\hat{\pi}_s = \frac{1}{2}(\hat{\pi}_s^A + \hat{\pi}_s^B) \tag{2.44}$$

If equal sample sizes were assumed (i.e., $n_{s1} = n_{s2} = n$), equal weights to the average of the estimators in (2.42) and (2.43), and defining $X^{K}_{sK,j+1} - X^{K}_{sK,j}$ to be the implied unobservable differences between the lists (Glynn, 2013), then the variance of $\hat{\pi}_s$ was:

$$Var(\hat{\pi}_s) = \frac{1}{4}\frac{1}{\frac{n}{2}}[V(X^{A}_{ns1} - X^{B}_{ns1})] + \frac{1}{4}\frac{1}{\frac{n}{2}}[V(X^{B}_{ns2} - X^{A}_{ns2})]$$

$$= \frac{1}{2n}\left\{V(X^{A}_{ns1}) + V(X^{B}_{ns1}) - 2Cov(X^{A}_{ns1}, X^{B}_{ns1})\right\}$$

$$+ \frac{1}{2n}\left\{V(X^{B}_{ns2}) + V(X^{A}_{ns2}) - 2Cov(X^{B}_{ns2}, X^{A}_{ns2})\right\}$$

$$= \frac{1}{2n}\{[V(X^{A}_{ns2}) + V(X^{A}_{ns1,j+1}) + 2Cov(X^{A}_{ns2}, X^{A}_{ns1,j+1})]$$

$$+ V(X^{B}_{ns1}) - 2[Cov(X^{A}_{ns2}, X^{B}_{ns1}) + Cov(X^{A}_{ns1,j+1}, X^{B}_{ns1})]\}$$

$$+ \frac{1}{2n}\{[V(X^{B}_{ns1}) + V(X^{B}_{ns2,j+1}) + 2Cov(X^{B}_{ns1}, X^{B}_{ns2,j+1})]$$

$$+ V(X^{A}_{ns2}) - 2[Cov(X^{B}_{ns1}, X^{A}_{ns2}) + Cov(X^{B}_{ns2,j+1}, X^{A}_{ns2})]\}$$

$$= \frac{1}{n}\{V(X^{A}_{ns2}) + V(X^{B}_{ns1}) + \frac{1}{2}[V(X^{A}_{ns1,j+1}) + V(X^{B}_{ns2,j+1})]$$

$$- 2Cov(X^{A}_{ns2}, X^{B}_{ns1})\} \tag{2.45}$$

$$+ \frac{1}{n}[Cov(X^{A}_{ns1,j+1}, X^{A}_{ns2}) - Cov(X^{B}_{ns2,j+1}, X^{A}_{ns2})] \tag{2.46}$$

$$+ \frac{1}{n}[Cov(X^{B}_{ns2,j+1}, X^{B}_{ns1}) - Cov(X^{A}_{ns1,j+1}, X^{B}_{ns1})] \tag{2.47}$$

If Assumption 1 (2.32) was met (i.e., no design effect) then the covariances in (2.46) and (2.47) were 0 and the variance of the estimator was (2.45). Note that this variance includes a correlation or covariance between the item lists. If in fact, the two item lists were highly correlated, variation of the estimator would be further reduced using the double item count technique (DICT; Glynn, 2013). By doing this, it is easy to see that the variance of the DICT (2.45) was more efficient then the variance of the item count technique (ICT; 2.38).

More recently, a fuzzy response model, the single sample count (SSC) technique was developed by Petróczi et al. (2011) in an attempt to simplify and provide a more economical form of the ICT. Like the relationship between the random response (RR) unrelated question technique (UQT) and FRT, in Petróczi's model, the need for an additional sample was eliminated by including innocuous questions in the item list with known probabilities of 0.50. The model was made more efficient since the need to estimate prevalence of the innocuous behaviors was eliminated and instead all survey participants would be used to estimate the sensitive behavior. Examples of innocuous questions used in the SSC would be "My birthday is in the first 6 months of the year," "My house number ends with an even number," and "The last digit of my telephone number is even." Since the number of innocuous questions in the item list would be known and have a 50-50 chance of endorsement, the estimator of the sensitive behavior could be easily derived by simply subtracting the proportion of endorsed items in the sample from the expected value of the endorsed items from the innocuous list of questions:

Let

$$X_{ij} = \begin{cases} 1 & \text{if the jth sample subject responds 'yes' to innocuous question i} \\ 0 & \text{if the jth sample subject responds 'no' to innocuous question i} \end{cases}$$

for the $i = 1.....m$ innocuous questions in the item list for $j = 1.....n$ sampled subjects.

and define

$$Y_j = \begin{cases} 1 & \text{if the jth sample subject responds 'yes' to the sensitive question} \\ 0 & \text{if the jth sample subject responds 'no' to the sensitive question} \end{cases}$$

for the same $j = 1.....n$ sampled subjects

Then

$$\frac{1}{2} = Pr[\text{jth sample subject responds yes to the innocuous question}]$$

$$\pi_s = Pr[\text{jth sample subject responds yes to the sensitive question}]$$

Define:

$$X = \sum_{j=1}^{n}(\sum_{i=1}^{m} X_{ij}) \tag{2.48}$$

$$Y = (\sum_{j=1}^{n} Y_j) \tag{2.49}$$

Then

$X : Bin(nm, 0.5)$

$Y : Bin(n, \pi_s)$.

Setting $n^{'}$ to nm and $p = \frac{1}{2}$ for X and $n^{'}$ to n and $p = \pi_s$ for Y, if $n^{`}p > 5$, $n^{'}(1-p) > 5$, and ($0.021 < \pi_s < 0.979$), then the general rules required to approximate a binomial distribution with a normal approximation can be made where:

$$X \approx N(\frac{1}{2}nm, \frac{1}{4}nm)$$

$$Y \approx N(n\pi_s, n\pi_s(1-\pi_s))$$

Let the sum of all "yes" responses in the sample be defined as

$\lambda = X + Y$, where

$$E[\lambda] = E[X] + E[Y]$$

$$= (nmp + n\pi_s).$$

Then the maximum likelihood estimator of π_s would be:

$$\hat{\pi}_s = \frac{\lambda}{n} - mp \qquad (2.50)$$

where

$$E[\hat{\pi}_s] = \frac{1}{n}E[\lambda] - mp$$

$$= \frac{1}{n}[(nmp + n\pi_s)] - mp$$

$$= mp + \pi_s - mp$$

$$= \pi_s.$$

Thus, $\hat{\pi}_s$ was an unbiased estimator of π_s.

The variance of $\hat{\pi}_s$ would be

$$Var[\hat{\pi}_s] = \frac{1}{n^2} Var[\lambda]$$

$$= \frac{1}{n^2}[(nmp(1-p) + n(\pi_s)1 - \pi_s]$$

$$= \frac{mp(1-p) + (\pi_s(1-\pi_s))}{n} \tag{2.51}$$

Since $\hat{\pi}_s$ was a maximum likelihood estimator, point estimates, and confidence intervals of the sensitive attribute can be made as usual.

Even though the utilization of the ICT and DICT has grown, the methodological research on the topic remained low. After a thorough review of the literature, there appeared to be no study--simulated or otherwise--comparing the efficiency of the technique to other techniques--such as randomized response or direct questioning. Recent methods that have improved efficiency of the technique included analysis by subpopulations or domains (Tsuchiya, 2005), modification in the manner in which the sensitive item was included (Chaudhuri & Christofides, 2007), adjustments to the difference in means estimator (Glynn, 2013) and the development of new nonlinear least squares and maximum likelihood estimators for multivariate analysis (Blair & Imai, 2012; Corstange, 2009; Imai, 2011). Since the SSC has been only recently developed, statistical efficiency between other techniques has not been formally explored.

Effects of Non-compliance in Random Response and Non-random Response Models

A review of the literature indicated that the primary focus of non-compliance in randomized response model has been to develop methods that estimated or adjusted for non-compliance resulting in an adjusted estimate of the sensitive behavior (Böckenholt et al., 2009; Böckenholt & van der Hejiden, 2004, 2007; Clark & Desharnais, 1998; Cruyff et al., 2007; Ostapczuk et al., 2010; van den Hout et al., 2010; van den Hout & Klugkist, 2009). In all cases, no assumptions were made regarding the intentions of the participants who are non-compliant. Clark and Densharais (1998) pioneered the efforts by extending the random response model to include estimates of the sensitive behavior prevalence and non-compliance when one question was used to measure the sensitive attribute. The model estimated three distinct population parameters: Honest yes (i.e., the proportion of compliant and honest "yes" participants, π_s), Honest no (i.e., the proportion of compliant and honest "no" participants, β) and SP (i.e., the proportion of non-compliant participants who respond negatively regardless of the randomized device, λ). A fourth proportion of participants, those who were non-compliant and responded "yes," were assumed negligible. The technique required two samples of subjects following the method described by Greenberg et al. (1969) where the selection of the random device for each sample was different but complementary. Using maximum likelihood estimation, closed form solutions of the parameters were provided. The authors also included a likelihood ratio test used to determine if the proportion of non-compliance was significant. In addition, power calculations used to determine optimal sample sizes that were meant to minimize non-compliance and thus reduce response bias were provided and discussed. In taking two separate samples, Clark and Densharais (1998) assumed that the level of non-compliant behavior was the same for each sample. In addition, using the equations provided by the authors, parameter estimates could fall outside the acceptable range of (0, 1). In order to improve on this limitation, using a medication non-adherence study, Ostapczuk et al. (2010) provided maximum likelihood estimates based on the more general family of multinomial models. Terming this the “cheating detection model” (CDM), the authors included an additional benefit of testing the significance of non-compliance in more complex models that include moderator variables. Using the CDM, Osapczuk provided estimates of the sensitive attribute based on the significance of the non-compliance. When non-compliance was not significant, asymptotic unbiased estimates of the sensitive attribute would be made; whereas when non-compliance was a significant factor, lower and upper bounds of the sensitive attribute would be made assuming the estimated proportion of non-compliant participants either all engaged or did not engage in the sensitive behavior.

Both Clark’s and Osapczuk’s models provided adjusted estimates of the sensitive behavior and non-compliance for random response models when only one sensitive question was used. Other methods were developed using several surveys from the Netherlands that included a series of randomized response questions--using a variety of randomized response techniques including Kuk’s (1990) method and the forced choice technique to estimate specific sensitive behaviors that included social security regulation infringements and social welfare fraud. Surveys were fielded in 2000, 2002, and 2004. Each survey contained a series of questions regarding the specific sensitive behaviors (i.e., social security fraud) that were ordered from less to more severe violations.

Böckenholt & van der Hejiden (2004, 2007) used the surveys and assumed an underlying non-compliant scale for the set of random response questions. They estimated non-compliance using an item-response model that incorporated a person level estimate based on Fox (2005). In their 2004 study, they distinguished between three classes of item response models, a model that assumed homogeneous compliance among participants, a model that allowed for individual variation of compliance between participants and a third model that also allowed for individual variation between participants but assumed a subset of non-compliant behavior. The authors developed techniques using maximum likelihood estimation and generalized ratio tests to show that the third model--adjusting for respondent variation and non-compliance--produced the best fit. Böckenholt & van der Hejiden (2007) extended this model to include mixture components: a component for individuals who followed the instructions of the randomized model and another component for individuals who were non-compliant. Using the 2002 and 2004 surveys, the authors concluded that mixture-item response models produced more accurate estimates of non-compliance and a better fit compared with item response models that did not account for non-compliance. Böckenholt & van der Hejiden (2007) used the 2000 survey to introduce a log linear model that provided adjusted estimates of the sensitive attribute prevalence as well as an estimate of non-compliance. They further developed this model by incorporating semi-parametric item response models that follow directly from latent class models (Böckenholt et al., 2009) using a dual design for direct questioning and the forced-choice randomized response technique. By adjusting for non-compliance and person level variables based on a series of questions measuring attitudes toward the sensitive domain, the authors were able to produce higher prevalence estimates of the sensitive behavior compared with models that did not adjust for these components. In addition, the model adjusting for non-compliance produced better fit statistics. By including person level attitude measures as well as non-compliance, the authors concluded, estimates of the sensitive attribute can be improved. van den Hout and Klugkist (2009) specified various models according to assumptions regarding non-compliance in an effort to extend the models introduced by Böckenholt & van der Hejiden (2007). Using the methods described in Rudas, Clogg, and Lindsay (1994), the authors used decreasing values of a goodness of fit test statistic and Bayesian inferences to determine the component weights. They extended the mixture component model to include Bayesian inference in estimating extended models and select between them. Non-compliance rates estimated using the Bayesian inferences were similar to those using the mixture component item-response models. As a result, the conclusion from the review of the literature in regards to non-compliance in random response model was to utilize methods that provided adjusted prevalence estimates of both the sensitive attribute and non-compliance. In general, these methods produced better fit statistics compared to the models that did not adjust for non-compliance. However there are no studies that demonstrated how different levels of non-compliance distorted prevalence estimates of the sensitive behavior especially between the UQT and FCT.

For the ICT, DICT, and SSC several studies have suggested that these techniques were not immune to non-compliant reporting and in effect, the techniques have produced estimates that appear to be under reported (Biemer & Wright, 2004; Droitcour et al., 1991; Holbrook & Krosnick, 2010; Kuklinski, Sniderman, et al., 1997). Tsuchiya et al. (2007) considered under-reporting a potential factor in the instability of estimates using

ICT. Under-reporting occurred when participants did not fully endorse the number of innocuous items for which they were members. As a result, the estimated prevalence of the sensitive attribute was under-estimated. In their study, Tsuchiya et al. (2007) determined that the prevalent rates of innocuous item lists questions asked in a direct questioning method were higher than corresponding rates from the same item lists asked in the item count technique (ICT) format. One reason highly cited in the literature for under-reporting in the ICT and DICT were the prevalence rates of each innocuous question in the item list (Blair & Imai, 2012; Corstange, 2009; Glynn, 2013). For the single sample count technique (SSC), this would not be an issue since prevalent rates of each innocuous question were approximately 0.50. However, literature regarding under-reporting in the SSC has not yet surfaced since the SSC was a new technique. Under-reporting was largely due to what is termed as “ceiling” or “floor” effects (Blair & Imai, 2012; Glynn, 2013; Tsuchiya et al., 2007). These effects occur when the item list of non-sensitive questions contained either a large proportion of highly prevalent items (ceiling effect) or an item list of non-sensitive questions with low prevalent items (floor effect). In the former case, since a high proportion of respondents possessed all characteristics on the item list, survey participants who possess the sensitive trait, could feel exposed and would therefore misreport their membership in the sensitive group. Likewise, for the latter case, since a high proportion of respondents possessed no characteristics of innocuous items, the survey participant with the sensitive attribute could again feel exposed, and under report their membership in the sensitive group. Blair and Imai (2012) demonstrated how ceiling and floor effects result in an under estimate of the sensitive attribute. This type of under-reporting membership in the sensitive group could be seen as non-complying. Kuklinski, Cobb, et al. (1997) had such a situation occur. In a study that examined non-southern attitudes toward racial profiling using the ICT, a large portion of the control sample reported “yes” to all of the non-sensitive questions. Since subjects were randomly assigned to receive the control or intervention list, similar patterns were assumed among intervention subjects. As a result, the estimate of the sensitive attribute, “racial profiling” was negative. This type of under reporting led Glynn (2013) to define three generally accepted practices regarding the design of the item count technique (ICT) and double item count technique (DICT) that were meant to lower under-reporting (i.e., non compliance):

1. Avoid large quantities of high prevalence non-sensitive items
2. Avoid large quantities of low prevalence non-sensitive items, and
3. Item lists should not be too short since shorter lists would increase the likelihood of ceiling effects and therefore under-reporting by the respondents with the sensitive attribute. However, at the same time item lists should not be made too long since longer lists increase variability

According to Glynn (2013) although increasing the list size could reduce bias (i.e., reduces the likelihood of ceiling and floor effects), at the same time model variability would also increase. This becomes the typical tradeoff between a bias (i.e., resulting likelihood of a ceiling/floor effects) estimate or an estimate with higher variability (increasing list size). In order to simultaneously minimize ceiling effects and response variability without compromising privacy, Glynn (2013) suggested a method defining an optimal design for the NRR. First, optimally allocating subjects into the two randomized group only minimally reduces variation (Glynn, 2013). Thus, Glynn (2013) suggested equally allocating subjects into the two groups. In fact, Glynn (2013) demonstrated how equal sample sizes would actually benefit the design--especially in terms of the double list technique. A more potential method in reducing variation, however, was the selection of the innocuous questions, their prevalence rates and how the items correlated. Glynn (2013) suggested selecting innocuous questions in the item list that correlate negatively. This would reduce the likelihood of ceiling and floor effects as well as variability since it decreased the number of “how many” items a respondent reported. For instance, two negatively correlated questions could be “I am not a pet owner” and “I shop at Petsmart”. Since the population of subjects who were not pet owners would most likely not be the same set of subjects who shopped at Petsmart, the number of “yes” responses would be reduced. Thus, for each technique (ICT, DICT, and SSC) by carefully correlating the questions in the innocuous item list, a researcher would effectively keep the list size relatively short and at the same time decrease the likelihood of ceiling or floor effects (Blair & Imai, 2012; Glynn, 2013). However, for the ICT, DICT, and SSC models with these optimal design features, a thorough examination of the effects of non-compliance on estimators has not been made. In addition, comparisons of statistical efficiency with other evasive response techniques--such as the RR--have also not been formerly studied.

Size Effects of Item Lists

For the ICT and SSC, the number of innocuous questions function as a cooperation variable much like the selection of *p* (the probability of selecting the sensitive question) in the RR models. If, for instance, a small number of innocuous questions were contained in the item list, there would be a higher likelihood of ceiling or floor effects (i.e., participant’s endorsing all/none of the questions) compared to a list with a larger number of innocuous questions. If this occurs, the participant’s membership to the sensitive group would be exposed and the purpose of the technique to increase cooperation would be compromised. However, at the same time, if the number of innocuous questions increased, not only would the survey become more burdensome to the participant, but the variation in the model would also increase. Of course ceiling and floor effects would be reduced by selecting non-sensitive questions in the item list that were highly negatively correlated (Blair & Imai, 2012; Corstange, 2009). The number of innocuous questions in an item list, therefore, were subject to both sampling (increased variation) and non-sampling error (Biemer & Wright, 2004; Tsuchiya et al., 2007). This would also be true for the single sample count technique (SSC) technique, which included a set of innocuous questions (Petróczi et al., 2011).

A review of the item count technique (ICT) literature suggests that for most models the optimal number of innocuous questions in the item list ranged between three and five (Ahart & Sackett 2004; Blair & Imai, 2012; Corstange, 2009; Dalton et al., 1994; Glynn, 2013; Tsuchiya et al., 2007; Wimbush & Dalton 1997). However, no empirical study, simulated or theorized, was made to determine how item list size influences estimates of varying sensitive prevalent rates (i.e., small, medium, and large) in the presence of both truthful and non-complying (i.e., under reporting) reporting. Only one empirical study comparing ICT to direct questioning in terms of shoplifting and blood donation rates, was made to determine the effects on prevalence, sampling and non-sampling error of ICT estimates compared with estimates using direct questioning (Tsuchiya et al., 2007). In their study Tsuchiya et al. (2007) compared estimates between a ranging number of item lists (two to five questions) using both direct questioning and ICT. Separate samples of participants from Japan were conveniently selected from a list of subjects and completed an online survey. Participants were randomly placed into 4 groups, where two of the groups completed the item lists in ICT format and two of the groups responded directly to each question in the item list. Item lists between each of the groups (i.e., the two ICT groups and the two direct question groups) alternated the sensitive question per list. The authors than compared the estimates between each sample by item list size to determine if item count technique (ICT) estimates were similar in estimating blood donation prevalence--a socially accepted behavior - and statistically significantly higher in estimating shop lifting prevalence--a socially stigmatizing behavior. The results indicated estimated rates of blood donation did not differ between the two methods with the exception of the 4-item list, where blood donation rates reported in the ICT were lower compared with the direct questioning (Tsuchiya et al., 2007). For shoplifting, the rates reported from the item list size of 2 and 4 using the ICT was statistically significantly higher compared to the corresponding item lists using the direct questioning technique (Tsuchiya et al., 2007). More importantly, even though the reported prevalent rates of shoplifting were higher from the ICT participants for item lists of size 4 and 5, the variation of these estimates was also higher. As a result, a statistically significant difference was not made between the two methods for these list sizes (Tsuchiya et al., 2007). Thus, the authors were able to demonstrate on an empirical level that although the ICT model may produce higher rate estimates compared to direct questioning, it also introduced a higher degree of statistical noise into the model.

Thus, an optimal number of innocuous questions used in an ICT, DICT or SSC item list was never formerly examined. If an optimal list would be determined, based on the prevalent rates of the sensitive attribute, this would provide, in theory, a set of guidelines experimenters can use that would in affect maximize honest reporting (i.e., reduce bias) and minimize variation.

CHAPTER III

METHODOLOGY

Background

This chapter has been broken down into three main parts: Part 1 includes a description of how the optimal design for the random response (RR) techniques (unrelated question [UQT] and forced choice [FCT]) and the non-random response (NRR) techniques (item count [ICT], double item count [DICT], and single sample count [SSC]) were determined. Part 2 describes the analytical methods used to evaluate the effects of non-compliance of and between these techniques. Part 3 describes the analytical methods used to evaluate the effect of list size for the ICT, DICT, and SSC.

Restating the research questions:

Q1 Are the indirect question techniques of the ICT, DICT, and SSC models more efficient, in the presence of non-compliant reporting, as measured by their Mean Squared Error (MSE) compared to the MSE of the RR models using the unrelated question technique and forced-choice techniques?

Q2 Is there an optimal number of innocuous questions in the item list for the ICT, DICT, and SSC techniques that will reduce non-compliance and minimize additional variation?

Selection of Optimal Design

Prior to testing the statistical efficiency between the two types of techniques (i.e., each RR vs. each NRR) in order to study non-compliance and effective list size, optimal design parameters of each technique as a function of the prevalence of the sensitive attribute (π_S) and sample size were found. Optimal designs were determined based on the selection of combined parameters from previous simulation studies regarding RR techniques (Greenberg et al., 1969; Lensvelt-Mulders, Hox, & Van der Heijden 2005) and ICT (Blair & Imai 2012; Corstange, 2009). Thus, a selected range of values of π_s similar to Lensvelt-Mulders, Hox, and van der Heijden (2005) arbitrarily categorized to represent small (i.e., 0.01, 0.03, and 0.05), medium (i.e., 0.10, 0.15, and 0.20), and large (i.e., 0.25, 0.35, and 0.45) prevalence rates in combination with sample sizes similar to the simulations performed by Blair and Imai (2012) and arbitrarily categorized as small (n = 150), medium (n = 500), and large (n = 1,500) were used. These same combinations of prevalence rates and sample sizes were also used in the simulation studies of non-compliance and list size.

Using Monte-Carlo simulations, the optimal design for each combination prevalent rate and sample size was selected and used in set of simulations studying the effects of non-compliance as well as studying the effects of list size. Thus, a total of 27 designs for each technique assessed non-compliance and list size. For the single sample count technique (SSC), $\pi_s = 0.01$ was eliminated from the analysis since normal approximations at this level could not be made (Petróczi et al., 2011). Therefore, 26 designs were used to assess non-compliance and list size for the SSC. Simulations selecting the optimal model were based on the efficiency study of Lensvelt-Mulders, Hox, and van der Heijden (2005). As was done in their study, truthful reporting (i.e., unbiased estimator) was assumed and since the direct question (DQT) technique was the most efficient (Greenberg et al., 1969; Lensvelt-Mulders, Hox, & van der Heijden, 2005), comparisons of the variance of each simulated technique to the variance of the direct questioning technique were made to determine the parameters of the optimal design (i.e., design parameters with the highest efficiency) for each of the 27 designs. Selecting optimal design parameters for each design within each technique were justified because it allows for an effective comparison of the effects of non-compliance by eliminating any additional source of variation due to the selection of the design parameters. Using the literature as a guide, the optimal design was determined by varying the design parameters for each technique.

For this set of simulations, the assumptions :

1. Truthful reporting (i.e., $\hat{\pi}_s$ is unbiased),
2. For the ICT, DICT, and SSC, no design effect, and
3. For the DICT, the 2-item lists were assumed correlated at 0.85, a correlation that is arbitrarily selected but practical since, in practice, lists can potentially be created with very high correlations. In addition, simulations of lists with the *between* list correlation using the "rmvbin" function in R produced valid results (i.e., all simulations resulted in an approximate multivariate normal distribution with a positive definite covariance matrix).

Relative reliability (RelRel) defined as the ratios of the variance of the DQT to the variance of the simulated technique (Kendall & Stuart, 1979; Lensvelt-Mulders, Hox, & van der Heijden, 2005) determined the most efficient design:

$$RelRel = \frac{\sigma_{sdqt}}{\sigma_{sim}} \qquad (3.1)$$

where the variance of the DQT was:

$$\sigma_{sdqt} = \frac{\pi_s(1-\pi_s)}{n} \qquad (3.2)$$

and σ_{sim} represents the variance of the comparing simulated technique (i.e., variance of the UQT, FCT, ICT, DICT, and SSC).

This type of analysis is known as a relative reliability study where variances between techniques are compared to determine statistical efficiency. The RelRel ratio is actually the inverse of the additional percentage of sampling units (in this case the unit would be survey participants) needed to obtain a variance comparable to the DQT technique. Thus if the sample size of both techniques were 100, a Rel Rel that is equal to 1/2 indicates that twice the number of participants (n = 200) from the comparing technique would be necessary to obtain a variance equal to the variance of the DQT; whereas a Rel Rel that is equal to 0.95 indicates that approximately the same number of sampling units (in this case increasing sample size of approximately five more participants) from the comparing technique would be necessary to obtain a variance equal to the variance of the DQT. In the first case, the comparing technique would be inefficient whereas in the second case, the comparing technique would be a more efficient estimator.

Random Response Techniques
Simulations

For the unrelated question technique (UQT), the design parameters simulated include p_1 (the probability of selecting the sensitive question in sample 1), p_2 (the probability of selecting the sensitive question in sample 2) as well as the prevalence rate of the innocuous question (π_{ns}). As discussed in Chapter II, the literature suggested that the value of p_1 and p_2 was most optimal when the design parameters are set as far apart, and on the opposite sides of 0.50 as possible, where $p_1 + p_2 = 1$ (Greenberg et al., 1969, Lensvelt-Mulders, Hox, & van der Heijden, 2005). This not only reduced the additional variation due to the introduction of the innocuous question but allowed for a symmetric and opposite effect of the probability of selecting the sensitive question in each sample (Greenberg et al., 1969). The literature also suggested that p_1 should range between 0.70

(+/- 0.10) and not exceed 0.80 (Greenberg et al., 1969; Lensvelt-Mulders, Hox, & van der Heijden, 2005). Since confidentiality was compromised when p_1 was too high or too low, simulations varied p_1 between 0.60-0.90 in increments of 0.10 (this, thus, varied p_2 between 0.10-0.40). These values were selected since they were similar to the simulations performed by Leansvelt-Mulder, Hox, and van der Heijden, (2005) which also encompassed the range of parameters simulated by Greenberg et al. (1969). For the prevalence rate of the innocuous question (π_{ns}), the literature suggested selecting an attribute with prevalence on the same side of 0.50 as the expected prevalent rate of the sensitive attribute (π_s) but large enough to ensure confidentiality (Greenberg et al., 1969; Lensvelt-Mulders, Hox, & van der Heijden 2005). Since the prevalence rates of the sensitive attributes simulated in this study were less than 0.50, similar parameters suggested by Greenberg et al. (1969) were used. Three values of π_{ns} were examined to arbitrarily represent small (π_{ns} = 0.10), medium (π_{ns} = 0.20) and large (π_{ns} = 0.30) prevalent rates. Although suggestions of these design parameters were first introduced by Greenberg et al. (1969), they have also been shown to produce more efficient estimates in other simulation studies (Lensvelt-Molders, Hox, & van der Heijden , 2005; Soeken & Macready, 1982). In addition, for UQT exclusively, sample sizes were allocated to each sample as described by Greenberg et al. (1969), where the proportion of the total sample size was allocated as:

$$\frac{n_1}{n_2} \equiv \sqrt{\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{\lambda_2(1-\lambda_2)(1-p_1)^2}} \, . \qquad (3.3)$$

What this formula does is to use the two components of the $\mathrm{var}(\hat{\pi}_s)$ (see equation 2.17) to ensure that a larger portion of subjects is allocated to the sample where the probability of selecting the sensitive question (i.e., p_1) is greater. For this method, since there are two varying design parameters, the method of Greenberg et al. (1969) was followed and simulations performed by first fixing p_1 and running simulations for each (π_{ns}). Since optimal sample sizes were based on p_1 and p_2, changes in sample size were subsequently adjusted as p_1 and p_2 change. Thus, the simulation design parameters, based on previous simulation studies, for selecting the optimal UQT were defined as follows:

1. The probability of selecting the sensitive question in sample 1 (p_1): 0.60, 0.70, 0.80, 0.90

2. The probability of selecting the sensitive question in sample 2 (p_2): 0.10, 0.20, 0.30, 0.40

3. The prevalence of the innocuous behavior (π_{ns}): 0.10, 0.20, 0.30

Since the FCT was the equivalent of the UQT, where the prevalence of the innocuous question was incorporated into the randomized device (Greenberg et al., 1969), a review of the literature determined that the most widely used design of this technique in examining non-compliance was the sum of two dice where p_1 (the probability of a forced "yes") was set at 1/6 (i.e., probability of observing a sum of 2, 3, or 4), p_2 (the probability of responding to the sensitive question) was set at 3/4 (i.e., probability of observing a sum between 5 and 10) and p_3 (the probability of a forced "no") was set at 1/12 (i.e., probability of observing a sum of 11 or 12; Böckenholt & van der Hejiden, 2004; van den Hout et al., 2010). This study followed these authors by using the FCT with these design parameters. Thus, for the FCT, the optimal model already been selected.

In the simulation study, for both random response (RR) techniques, likelihood functions defined in Chapter II were used to derive the variance of each set of parameters based on sample size and π_s. These were calculated exactly and compared against the corresponding variance of the DQT. For completeness, the variance of the FCT will also be compared.

Indirect Question Techniques
Simulations

For the set of simulations regarding the indirect question techniques, the study assumed no design effect. That is the introduction of the sensitive question into the item list were assumed not to change the nature of responses to the non-sensitive questions in the item list.

For the indirect question techniques (ICT and SSC), design parameters were determined by list size of the non-sensitive questions and included the correlation between these items in each list as well as the distribution of prevalence rates of each innocuous question in the list. For the DICT, an initial study of design parameters by list size of the non-sensitive questions included correlations between questions *within* a list and correlations of all questions *between* lists by performing simulations that used the "rmvbin" function in R. The "rmvbin" function was selected since it simulates multivariate binary distributions with specified correlations (Leisch, Weingessel, & Hornik, 1998) that draw samples from a corresponding multivariate normal distribution, with the same mean and covariance structure. Using this distribution, by thresholding, the algorithm converted multivariate normal data into multivariate binary data with the appropriate mean and covariance structure. For a more detailed explanation, see Leisch et al. (1998). Recall that the double item count technique (DICT) contained two lists of questions, where list A contained a set of innocuous questions and list B contained a different set of innocuous questions. For each sample of participants, the first sample responded to list A with the inclusion of the sensitive question as well as list B, which only included the set of innocuous questions. The second sample then responded to list A which included just the set of innocuous questions and list B which included the additional sensitive question. *Within* list correlations occurred when two questions within the *same* list were negatively correlated. So for instance, if the list size were 3, 2 questions (i.e., say, innocuous question 1 and innocuous question 2 in list A) *within* the item list were correlated. *Between* list correlations occurred when all innocuous questions *between* lists were positively correlated. So for instance in a 3-item list, a between list correlation would be positively correlating innocuous question 1 of each list, innocuous question 2 of each list and innocuous question 3 of each list. This, according to Glynn (2013), would reduce variation in the model. When both *within* and *between* correlations were attempted in simulations, the multivariate normal distribution used to simulate the list of binary data resulted in a non-positive definite covariance matrix for all list size, sample size and sensitive prevalent rates. Because of this, a second set of simulations were performed to determine if DICT models were more efficient when correlations *within* list items were made or correlations *between* list items were made. These sets of simulations are discussed in Chapter IV under the section detailing the analysis and results of the selection of the DICT optimal model. In short, the study indicated that the model with *between* list correlations was the most efficient. Because of this, for the simulations determining the optimal DICT model, no *within* list correlations were made. Instead only *between* list correlations were simulated (see Chapter IV for a discussion of the results).

According to the literature, list sizes usually ranged between three and five innocuous questions (Blair & Imai, 2012; Glynn, 2013; Tsuchiya et al., 2007). For this study, list sizes between three (small) and five (large) which were similar to the Tsuchiya et al. (2007) study were simulated. Since the literature suggested correlating questions in the list of innocuous questions not only reduced ceiling and floor effects (Blair & Imai, 2012; Corstange, 2009; Glynn, 2013), but also variation (Blair & Imai, 2012; Corstange, 2009; Glynn, 2013), negatively correlating pairs of non-sensitive questions *within* an item list were considered for the item count technique (ICT) and single sample count technique (SSC); whereas positive correlations *between* item lists were considered for the double item count technique (DICT). In addition, in his discussion of optimal design parameters for the ICT and DICT, Glynn (2013) suggested that ceiling and floor effects were further reduced if sensitive questions with high prevalence and low prevalence are avoided in abundance. Thus, this study chose to follow and expand on the simulation studies of Blair and Imai (2012) and Corstange (2009) who simulated distributions of the prevalent rate of innocuous questions based as “equal,” “not equal and symmetric,” and “not equal and not symmetric.” Each of the prevalent rate distributions, therefore, followed the suggestion of Glynn (2013) and were meant to simulate lists of non-sensitive questions that control for ceiling and floor effects by reducing non-sensitive questions within the list with high and/or low prevalent rates in abundance. By doing this, conclusions regarding the distribution of the prevalent rates of the non-sensitive questions and statistical efficiency were studied. Distributions of the non-sensitive item prevalent rates were determined based on list size (3-item, 4-item, and 5-item) for the item count technique (ICT) and double item count technique (DICT). Since prevalent rates of the single sample count technique (SSC) are 0.50 for all non-sensitive questions, this was not an issue, and the distribution for this technique was assumed "equal." Because the study considers specific prevalent rates of non-sensitive items (i.e., "equal," "not equal and symmetric," and "not equal not symmetric"), in order to negatively correlate pairs of these items for the ICT and SSC the correlation selected must be valid given the marginal probabilities (i.e., prevalent rates of the non-sensitive questions to be correlated). In the case of two random variables, allowable correlations were restricted by the joint probability, which is bounded by the minimum marginal probability (Leisch et al., 1998). For instance, it is not possible to simulate two correlated binary variables at $\rho = -0.50$ with marginal probabilities of 1/5 and 2/5 since the resulting joint probability would be -0.018, which is not a valid probability. In the case of this simulation, the marginal probabilities represented the prevalent rates of the non-sensitive and sensitive questions in the item list. Since several distributions were explored in this study, a *within* list correlation of -0.50 and a *between* list correlation of 0.85 were selected since all selected correlated pairs of non-sensitive items resulted in valid joint probabilities.

List sizes were simulated using Monte Carlo simulations by generating a series of Bernoulli random variables in the R program using the function, "rmvbin" for each of the three sample sizes (i.e., 150, 500, and 1,500). Correlations occurred in sequences of pairs based on the probability distribution of the innocuous questions ("equal," "not equal and symmetric," and "not equal and not symmetric") and allowable correlation of -0.50

between pairs of sensitive questions. In addition, a Bernoulli random variable was also simulated to represent the sensitive question. This was based on the prevalent rate of the sensitive attribute being simulated (π_s) where honest reporting was assumed. Since sample size does not substantially improve efficiency in the ICT and double item count technique (DICT; Blair & Imai, 2012; Glynn, 2013), sample sizes were allocated equally to each group (i.e., simulated group receiving the sensitive question, simulated group not receiving the sensitive question).

The number of simulations performed was determined by a preliminary simulation study that ensured estimation of the variability of the variance within 0.01. Once this number was decided, simulations were made for each sample size and prevalent rate combination and each variation of list size (3 item, 4-item, and 5-item), correlation *within* a list (i.e., 0, -0.50 for ICT and SSC only) or *between* a list (0.85) and prevalence rate of the non-sensitive questions in the item list ("equal," "not equal but symmetric," and "not equal and not symmetric"). For each set of simulations, difference-in-means and average difference-in-means estimates were obtained for the ICT and DICT, respectively, as well as estimators of SSC. Variances of these estimates were then compared with the variance of DQT to determine efficiency. Optimal models were found by list size, meaning there were a total of 27 optimal models (i.e., by sample size and prevalent rate combination) for each list size, with the exception of the SSC in which case 26 optimal models were determined.

Simulation parameters for the ICT were defined as:

1. Item List size: (3, 4, and 5)
2. Correlation *within* the item list: (0.0, -0.50)

3. Probabilities of the non-sensitive items:

    a. 3-item list (* indicate pairs of non-sensitive items to be correlated):

        i. Equal: (2/3, *2/3, and *2/3),

        ii. Not equal but symmetric: (1/4, *1/2, and *3/4)

        iii. Not equal and not symmetric: (1/4, *2/3, and *2/3)

    b. 4-item list (* and + indicate pairs of non-sensitive items to be correlated)

        i. Equal: (*2/3, *2/3, +2/3, and +2/3)

        ii. Not equal but symmetric: (1/5, 2/5, *3/5, and *4/5)

        iii. Not equal and not symmetric: (1/6, 3/6, *4/6, and *4/6)

    c. 5-item list (* and + indicate pairs of non-sensitive items to be correlated)

        i. Equal: (*2/3, *2/3, +2/3, +2/3, and 2/3)

        ii. Not equal but symmetric: (1/6, *2/6, +3/6, +4/6, and *5/6)

        iii. Not equal and not symmetric: (1/7, *3/7, +4/7, +5/7, and *5/7)

Simulation parameters for the DICT were defined as:

1. Item List size: (3, 4, and 5)

2. Correlation between lists: 0.85

3. Probabilities of the non-sensitive items:

    a. 3-item list:

        i. Equal: (2/3, 2/3, and 2/3),

        ii. Not equal but symmetric: (1/4, 1/2, and 3/4)

iii. Not equal and not symmetric: (1/4, 2/3, and 2/3)

b. 4-item list:

i. Equal: (2/3, 2/3, 2/3, and 2/3)

ii. Not equal but symmetric: (1/5, 2/5, 3/5, and 4/5)

iii. Not equal and not symmetric: (1/6, 3/6, 4/6, and 4/6)

c. 5-item list:

i. Equal: (2/3, 2/3, 2/3, 2/3, and 2/3)

ii. Not equal but symmetric: (1/6, 2/6, 3/6, 4/6, and 5/6)

iii. Not equal and not symmetric: (1/7, 3/7, 4/7, 5/7, and 5/7)

Simulation parameters for the SSC were defined as:

1. Item List size: (3, 4, and 5)

2. Correlation *within* the item list: (0.0, -0.50; * and $^{+}$ indicate pairs of non-sensitive items to be correlated)

   a. 3-item List (1/2, *1/2, and *1/2),

   b. 4-item List (*1/2, *1/2, $^{+}$1/2, and $^{+}$1/2)

   c. 5-item List (*1/2, *1/2, $^{+}$1/2, $^{+}$1/2, and 1/2)

Studying the Effects of Non-compliance

In order to examine Research Question 1, Are the indirect question techniques of the ICT, DICT, and SSC models more efficient as measured by their Mean Squared Error (MSE) to non-compliant responding compared to the MSE of the RR models using the unrelated and forced-choice techniques, a second set of simulations were performed. For this study, examination of the effects of non-compliance based on the same range of values of π_s that were previously used were made: small (i.e., 0.01, 0.03, and 0.05), medium (i.e., 0.10, 0.15, and 0.20), and large (i.e., 0.25, 0.35, and 0.45) in combination with small (*n* = 150), medium (*n* = 500), and large (*n* = 1,500) sample sizes. However, in this set of simulations, the optimal design parameters of each technique for each π_s, sample size combination from the previous simulation study were used.

For this study, the simulation study of Greenberg et al. (1969) was followed, where compliance (T) was defined in terms of the percent of participants who responded truthfully and ranged between T = 1.00 (all respond truthfully) and T = 0.40 (only 40% of respondents answer truthfully) in increments of 0.10. This corresponds to non-compliance rates ranging between 0.00 (no misreports) to 0.60 (60% of sensitive population misreports). Non-compliance was also defined as was done in the study of Greenberg et al. (1969), where non-compliance only occurred among participants with the sensitive attribute who were asked the sensitive question. Thus, non-complying responses for each technique were defined as:

1. Unrelated question technique (UQT): Both sets of participants (i.e., those with and without the sensitive trait) responded truthfully when asked the innocuous question, participants without the sensitive trait responded truthfully to the sensitive question (i.e., respond “no”). Misreporting will only occurred among participants with the sensitive trait when asked the sensitive question.

2. Forced choice technique (FCT): this assumed that both sets of participants (i.e., those with and without the sensitive trait) responded truthfully when forced to say “no” (i.e., sum of the dice is 11 or 12) or “yes” (i.e., sum of dice is between 2 and 4). Misreporting occurred among participants with the sensitive trait who are asked the sensitive question (i.e., sum of dice was between 5 and 10).

3. For the indirect question techniques (ICT, DICT, and SSC), the study assumed that both sets of participants (i.e., those with and without the sensitive trait) responded truthfully to the list of non-sensitive questions and that participants without the sensitive trait responded truthfully to the sensitive question (i.e., respond "no"). Misreporting only occurred among participants with the sensitive trait, in which case the "yes" responses to the list of questions were under-reported by 1.

Using the optimal design parameters of each technique, for each of the 27 sample size sensitive prevalent rate (π_s) designs, non-compliance ranging from 0 (truthful) to 0.40, in increments of 10 was examined. In each case, the ratio of the mean squared error (MSE) was calculated and used to compare the effects of non-compliance between techniques. For the random response (RR) techniques (UQT and FCT), maximum likelihood estimators were calculated directly. For the indirect question techniques (ICT, DICT, and SSC), simulations were made as was described previously for the selection of the optimal design. In this set of simulations, however, optimal design parameters for each list size (3-item, 4-item, and 5-item) were used. In order to account for the rate of non-compliance, using the "rmvbin" function in R, the simulated Bernoulli random variable meant to represent the prevalent rate of the sensitive question was based on the proportion of those expected to report truthfully (i.e., π_s to $T\pi_s$). Variance and bias of these estimates were then used to calculate the MSE of each technique.

Effects of non-compliance between techniques were made by comparing the ratio of MSEs between each of the RR techniques to each of the item count technique (ICT) techniques given the sample size (150, 500, and 1,500), prevalence of the sensitive attribute (π_s), percentage of truthful reporting (0.40-1) by list size (3-item, 4-item, and 5-

item) of the ICT, DICT, and SSC. So for instance to determine the effects of non-compliance between techniques given a small sample size (i.e., n = 150), small sensitive prevalent rate (π_{ns} = 0.03), and a small percentage of non-compliance (i.e., T = 0.90), using optimal model parameters from simulation study 1, the MSE for each random response (RR) technique were compared to the MSE of each non-random response (NRR) technique by list size:

$$Ratio(MSE) = \frac{MSE_{rr}}{MSE_{iqt}}, \tag{3.4}$$

where

$$MSE_{rr} = Var(\hat{\pi}_s) + Bias(\hat{\pi}_s)^2, \tag{3.5}$$

$$MSE_{iqt} = Var(\hat{\pi}_s) + Bias(\hat{\pi}_s)^2. \tag{3.6}$$

As a result, based on sample size and expected sensitive prevalent rate, the technique that performs more efficiently in the presence of non-compliance was determined.

Studying the Effects of List Size in the Item Count Technique, Double Item Count Technique, and Single Sample Count

In order to examine Research Question 2, Is there an optimal number of innocuous questions in the item list for the item count technique (ICT) and single sample count technique (SSC) techniques that will reduce non-compliance and minimize additional variation?, the results of the previous two simulation studies were used to examine the effects of list size on statistical efficiency for each NRR technique by non-compliance rate. For this study, the effects of list size were examined based on the same range of values of π_s that were previously used: small (i.e., 0.01, 0.03, and 0.05), medium (i.e., 0.10, 0.15, and 0.20), and large (i.e., 0.25, 0.35, and 0.45) in combination with small (*n* = 150), medium (*n* = 500), and large (*n* = 1,500) sample sizes, with the exception of the SSC in which π_s = 0.01 was eliminated. However, no further simulations were made. Instead comparison of the ratio of the MSEs found in simulation study 2 within each technique (i.e., ICT, DICT, SSC) by sample size (150, 500, and 1,500), sensitive prevalent rate (0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, and 0.45) and percent of non-compliance (1.00-0.40 in increments of 10) across list size (3-item, 4-item, and 5-item) were used. Comparisons were made between all combinations of list sizes. So for instance, ratios of MSE for the ICT using the optimal design parameters of π_s = 0.03 and complaint rate of T (i.e., percent of truthful reporting) = 0.90 were compared by all combinations of list size (3-item, 4-item, and 5-item). By examining the efficiency within each NRR technique by list size, the study was able to determine if smaller list sizes were just as efficient as larger list sizes in the presence of both truthful responding (simulation 1) and non-compliance (simulation 2).

In addition, efficiency between NRR techniques was also explored by comparing the ratios of the MSE found in simulation studies 1 and 2 by sample size (150, 500, and 1,500), sensitive prevalent rate (0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, and 0.45), percent of non-compliance (1-0.40 in increments of 10) and list size (3-item, 4-item, and 5-item) *between* NRR techniques. So for instance, in order to determine if the DICT was more efficient than the ICT for a list size of 3 when π_s = 0.03, *n* = 150, and compliance is T = 0.90, the ratios of the MSEs between these techniques at this list size, using optimal design parameters, were analyzed descriptively. By examining the efficiency between each NRR technique by list size, the study was able to determine if certain NRR techniques were more efficient in the presence of both truthful responding (simulation 1) and non-compliance (simulation 2).

CHAPTER IV

RESULTS

Background

This chapter has been broken down into four main parts. Part 1 includes a description of how the total number of simulations was determined for the study. Part 2 describes the results of selecting the optimal design for the random response (RR) techniques (unrelated question [UQT], forced choice [FCT]) and the non-random response (NRR) techniques (item count [ICT], double item count [DICT]. and single sample count [SSC]). Part 3 describes the results of the study of non-compliance between RR techniques and NRR techniques. Finally, Part 4 describes the results for the effects of list size for the NRR techniques.

Determining Number of Simulations

Prior to running final simulations for item count technique (ICT), double item count technique (DICT), and single sample count technique (SSC) in determining optimal models as well as the effects of non-compliance and list size, a preliminary study was made to determine an efficient number of simulations to perform. In Chapter III, it was stated that the total number of simulations run would ensure estimation of the variability of the variance within 0.01. To determine this number, a simulation study of the variance based on the technique with the highest variability was made since it ensured techniques with less variability would also be within this defined bound. Since the literature suggested that the DICT produced a more efficient estimator compared with the estimator of the item count technique (ICT; Glynn, 2013; Tsyuchiya et al, 2007), it was decided to examine the variability between the 5-item ICT and 5-item SSC procedures to determine which of these models produced the greatest amount of model variability. For this set of simulations, the models with the expected highest variability would be those with the smallest sample sizes (*n* = 150), no correlation between sensitive questions and with the largest sensitive prevalence rate ($\pi_S = 0.45$). Each of these attributes contributed to higher model variability since smaller samples sizes generally increase variation, and, if the estimated sensitive attribute is large, this in turn increased the likelihood that the range of sums across each list would be wider (i.e., more likely to have sums range between 0 and 6 than if the sensitive prevalence was smaller)--which would contribute to higher variability. In addition, models that do not take advantage of purposely correlating non-sensitive items also are expected to result in higher variability (Glynn, 2013).

For this set of simulations, an arbitrary number of 30 sets of 500 simulations were made for each of the 5-item list of ICT models by the distribution of the non-sensitive items: "equal," "not equal and symmetric," and "not equal and not symmetric." In addition, 30 sets of 500 simulations were run for the 5-item list of SSC models. As a result, a total of four models were simulated. For each simulation, item lists of 75 per sample for the ICT and 150 per sample for the SSC were generated using the "rmvbin" function in R. The difference-in-means estimator and the expected value estimator were then calculated for the ICT and SSC, respectively. After simulating 500 sets of estimators, the variance was calculated and used as an estimate. This was repeated 30 times for each model resulting in 30 variances of each estimator. Using these 30 variances, descriptive statistics including the mean, median, minimum, maximum, and variance were studied and compared between models. Table 1 displays the results of the simulation study. Results indicated that the 5-item ICT with “equally” distributed non-sensitive item prevalence rates had the highest variation compared with all other models and the 5-item ICT with a distribution of “not equal and not symmetric” prevalence rates had the highest variability of variances (Table 1).

Table 1

Simulation Study of Variability Between 5-Item Count Technique and 5-Item Single Sample Count to Determine Total Number of Study Simulations Performed

Descriptive	Item Count Technique: Variances			Single Sample Count: Variance
	Equal	Not Equal & Symm	Not Equal & Not Symm	
Minimum	0.0307	0.0259	0.0273	0.0088
Median	0.0329	0.0289	0.0303	0.0100
Mean	0.0331	0.0288	0.0304	0.0100
Maximum	0.0370	0.0316	0.0348	0.0113
Variance	**2.5613E-06**	2.1742E-06	**3.6404E-06**	2.9485E-07

Note. $\pi_s = 0.45$, $n = 150$

As a result, both the 5-item ICT with “equal” non-sensitive question prevalence rates (2/3, 2/3, 2/3, 2/3, and 2/3) and the 5-item ICT with “not equal and not symmetric” non-sensitive question prevalence rates were used in the next set of simulations. Using the literature as a guide (Blair & Imai, 2012; Cornstange, 2009), it was decided to test the effect of 1,000 simulations on the variability of both models. Simulations were performed similarly to the previous simulation, where 100 sets of 1,000 5-item lists were generated using the "rmvbin" function in R for each model. For each of the 100 sets, a total of 1,000 difference-in-means estimators were calculated and the variance of these simulated estimators taken. This resulted in a total of 100 variances of the ICT in which the variance of these variances were studied. For the ICT with "equal" non-sensitive prevalence rates, the variance of variances was 0.00000279, whereas for the ICT with "not equal and not symmetric" non-sensitive prevalence rates resulted in a variance of the variances of 0.00000196. Both sets of simulation resulted in variances well below the 0.01 bound. Thus, for this study, it was determined 1,000 simulations would result in a good estimate of the variability in simulations for the ICT, DICT, and SSC.

Determining the Optimal Model

Unrelated Question Technique (UQT)

Table 2 displays the results from the study determining the optimal model parameters for the UQT, optimal models for each sensitive prevalent rate and sample size are bolded in the table. As can be seen in Table 2, the model producing the highest relative reliability under the assumption of truthful reporting was the same across all sensitive prevalent rates and sample sizes. This model was the one adjusting for the lowest non-sensitive prevalence rate (π_{ns} = 0.10), where the probability of selecting the sensitive question in the first sample was the farthest (p_1 = 0.90) from the complementary probability of selecting the sensitive question for the second sample (p_2 = 0.10). This was not surprising since Greenberg et al. (1969) indicated that reduction of variation in the UQT occurred when model parameters were as far from 0.50 as possible. In fact, Greenberg et al. (1969) indicated that, for the UQT, variances could be made as close as possible to the binomial if the probability of selecting the sensitive questions for each sample were as far apart from one another--and 0.50--as allowable. Likewise, the prevalence rate of the innocuous question should also be as far from 0.50 as possible since this too reduces variation (Greenberg, 1969). Table 2 also demonstrated that, when comparing the unrelated question technique (UQT) variability to the variability of the DQT, sample size was not a factor. This was due to the fact that since sample sizes were equal, when taking the ratio of the MSEs using the likelihood function, the sample size cancels out. In order to study the effects of the UQT model parameters on variation, Figure 1 displays the relative reliability ratios by the UQT parameters for the sensitive prevalence rates, $\pi_s = 0.05$, $\pi_s = 0.20$, and $\pi_s = 0.45$, when $n = 500$. These rates were selected to represent sensitive prevalent rates that are small, medium and large. In examining the figure, it was clear that, by changing the parameter p_1 (the probability of selecting the sensitive question in sample 1 and likewise sample 2), the variability of the UQT increased (i.e., relative reliability decreases) more substantially compared to the change in variability due to the parameter π_{ns}.

Since the optimal model of the UQT selected in this study was not the most practical model--since according to Greenberg et al. (1969)--the probability of selecting the sensitive question in sample 1 was so high and the prevalence rate of the non-sensitive question so low ($\pi_{ns} = 0.10$), confidentiality may be comprised if this model were used in practice. Since the UQT model with the lowest relative reliability ($p_1 = 0.60$, $p_2 = 0.40$ and $\pi_{ns} = 0.30$) was the model most likely to increase confidentiality, this model was also explored in the simulation study of non-compliance.

Table 2

Unrelated Question Technique Relative Reliability, Selection of the Optimal Model

π_s	π_{ns}	p_1	Relative Reliability n = 150	Relative Reliability n = 500	Relative Reliability n = 1,500
0.01	0.10	0.9	**0.2756**	**0.2756**	**0.2756**
		0.8	0.1021	0.1021	0.1021
		0.7	0.0359	0.0359	0.0359
		0.6	0.0079	0.0079	0.0079
	0.20	0.9	0.1764	0.1764	0.1764
		0.8	0.0595	0.0595	0.0595
		0.7	0.0202	0.0202	0.0202
		0.6	0.0044	0.0044	0.0044
	0.30	0.9	0.1325	0.1325	0.1325
		0.8	0.0433	0.0433	0.0433
		0.7	0.0145	0.0145	0.0145
		0.6	0.0032	0.0032	0.0032
0.03	0.10	0.9	**0.4706**	**0.4706**	**0.4706**
		0.8	0.2162	0.2162	0.2162
		0.7	0.0844	0.0844	0.0844
		0.6	0.0196	0.0196	0.0196
	0.20	0.9	0.3547	0.3547	0.3547
		0.8	0.1435	0.1435	0.1435
		0.7	0.0525	0.0525	0.0525
		0.6	0.0118	0.0118	0.0118
	0.30	0.9	0.2905	0.2905	0.2905
		0.8	0.1104	0.1104	0.1104
		0.7	0.0394	0.0394	0.0394
		0.6	0.0088	0.0088	0.0088
0.05	0.10	0.9	**0.5528**	**0.5528**	**0.5528**
		0.8	0.2798	0.2798	0.2798
		0.7	0.1158	0.1158	0.1158
		0.6	0.0278	0.0278	0.0278
	0.20	0.9	0.4473	0.4473	0.4473
		0.8	0.2003	0.2003	0.2003
		0.7	0.0772	0.0772	0.0772
		0.6	0.0178	0.0178	0.0178
	0.30	0.9	0.3831	0.3831	0.3831
		0.8	0.1603	0.1603	0.1603
		0.7	0.0597	0.0597	0.0597
		0.6	0.0136	0.0136	0.0136

Table 2 (continued)

π_s	π_{ns}	p_1	Relative Reliability n = 150	Relative Reliability n = 500	Relative Reliability n = 1,500
0.10	0.10	0.9	**0.6400**	**0.6400**	**0.6400**
		0.8	0.3600	0.3600	0.3600
		0.7	0.1600	0.1600	0.1600
		0.6	0.0400	0.0400	0.0400
	0.20	0.9	0.5596	0.5596	0.5596
		0.8	0.2856	0.2856	0.2856
		0.7	0.1188	0.1188	0.1188
		0.6	0.0286	0.0286	0.0286
	0.30	0.9	0.5061	0.5061	0.5061
		0.8	0.2428	0.2428	0.2428
		0.7	0.0973	0.0973	0.0973
		0.6	0.0229	0.0229	0.0229
0.15	0.10	0.9	**0.6755**	**0.6756**	**0.6756**
		0.8	0.3966	0.3967	0.3967
		0.7	0.1819	0.1819	0.1819
		0.6	0.0463	0.0463	0.0463
	0.20	0.9	0.6114	0.6114	0.6114
		0.8	0.3321	0.3321	0.3321
		0.7	0.1440	0.1441	0.1441
		0.6	0.0355	0.0355	0.0355
	0.30	0.9	0.5675	0.5675	0.5675
		0.8	0.2925	0.2925	0.2925
		0.7	0.1225	0.1225	0.1225
		0.6	0.0296	0.0296	0.0296
0.20	0.10	0.9	**0.6931**	**0.6931**	**0.6931**
		0.8	0.4155	0.4155	0.4155
		0.7	0.1934	0.1934	0.1934
		0.6	0.0497	0.0497	0.0497
	0.20	0.9	0.6400	0.6400	0.6400
		0.8	0.3600	0.3600	0.3600
		0.7	0.1600	0.1600	0.1600
		0.6	0.0400	0.0400	0.0400
	0.30	0.9	0.6033	0.6034	0.6034
		0.8	0.3247	0.3247	0.3247
		0.7	0.1399	0.1399	0.1399
		0.6	0.0343	0.0343	0.0343

Table 2 (continued)

π_s	π_{ns}	p_1	Relative Reliability n = 150	Relative Reliability n = 500	Relative Reliability n = 1,500
0.25	0.10	0.9	**0.7017**	**0.7018**	**0.7018**
		0.8	0.4247	0.4247	0.4247
		0.7	0.1990	0.1991	0.1991
		0.6	0.0514	0.0514	0.0514
	0.20	0.9	0.6568	0.6568	0.6568
		0.8	0.3770	0.3770	0.3770
		0.7	0.1700	0.1700	0.1700
		0.6	0.0429	0.0429	0.0429
	0.30	0.9	0.6258	0.6259	0.6259
		0.8	0.3460	0.3460	0.3460
		0.7	0.1519	0.1519	0.1519
		0.6	0.0377	0.0377	0.0377
0.35	0.10	0.9	**0.7045**	**0.7045**	**0.7045**
		0.8	0.4268	0.4269	0.4269
		0.7	0.2001	0.2001	0.2001
		0.6	0.0516	0.0516	0.0516
	0.20	0.9	0.6708	0.6708	0.6708
		0.8	0.3915	0.3916	0.3916
		0.7	0.1787	0.1787	0.1787
		0.6	0.0454	0.0454	0.0454
	0.30	0.9	0.6483	0.6484	0.6484
		0.8	0.3684	0.3685	0.3685
		0.7	0.1649	0.1650	0.1650
		0.6	0.0414	0.0414	0.0414
0.45	0.10	0.9	**0.6943**	**0.6944**	**0.6944**
		0.8	0.4148	0.4149	0.4149
		0.7	0.1922	0.1922	0.1922
		0.6	0.0493	0.0493	0.0493
	0.20	0.9	0.6688	0.6688	0.6688
		0.8	0.3891	0.3891	0.3891
		0.7	0.1771	0.1771	0.1771
		0.6	0.0449	0.0449	0.0449
	0.30	0.9	0.6527	0.6527	0.6527
		0.8	0.3728	0.3728	0.3728
		0.7	0.1675	0.1675	0.1675
		0.6	0.0421	0.0421	0.0421

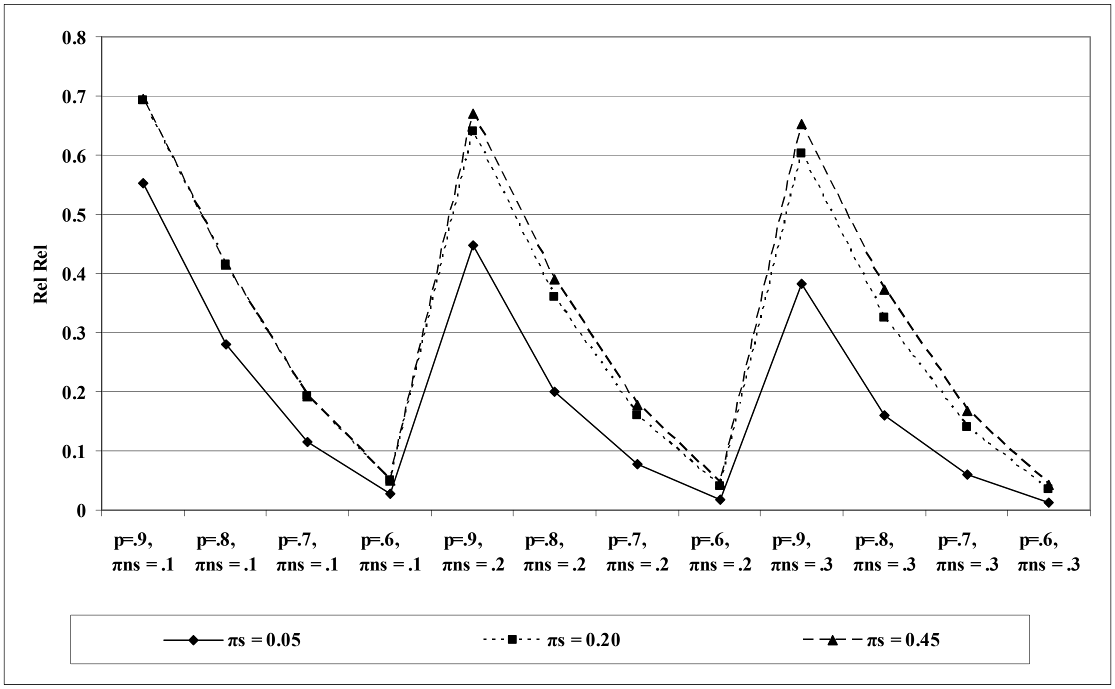


Figure 1. Relative Reliability, Direct Questioning Technique vs. Unrelated Question Technique under the assumption of truthful reporting, n = 500

Forced Choice Technique (FCT)

As was stated in Chapter III, a review of the literature determined that the most widely used FCT design in examining non-compliance was the sum of two dice where p_1 (the probability of a forced "yes") was set at 1/6 (i.e., probability of observing a sum of 2, 3, or 4), p_2 (the probability of responding to the sensitive question) was set at 3/4 (i.e., probability of observing a sum between 5 and 10), and p_3 (the probability of a forced "no") was set at 1/12 (i.e., probability of observing a sum of 11 or 12; Böckenholt et al., 2009; Böckenholt & van der Heijden, 2004, van den Hout et al., 2010). This study chose to follow these authors and use the FCT with these design parameters. Table 3 displays the relative reliability results of the FCT as compared with the DQT under the assumption of truthful reporting. As is evident in the table, the relative reliability was the same across sample sizes. This was due to the fact that since sample sizes were equal, when taking the ratio of the MSEs using the likelihood function, the sample size cancels out.

As is evident in Table 3, as the sensitive prevalent rate increased from very small ($\pi_S = 0.01$) to very large ($\pi_S = 0.45$) so to did the relative reliability. This was primarily due to the fact that the variance of the DQT was smallest when the prevalence rate was very small and approached maximum as the prevalence rate increased toward 0.50.

Item Count Technique (ICT)

For the item count technique (ICT), two sets of simulations were run. In the first set, simulations of non-sensitive questions within the item list were not purposely correlated and in the second set, specific non-sensitive questions were correlated at -0.50 as described in Chapter III. Simulations were then performed for each prevalent rate - sample size combination, where the distribution of prevalent rates of the non-sensitive items were either “equal,” “not equal but symmetric,” and “not equal and not symmetric,” resulting in a total of 81 simulations per list size. For each simulation, 1,000 pairs of item lists were created using the R function, “rmvbin,” resulting in a total of 1,000 difference-in-means estimates for which the variance was calculated and used as an estimate of the variability in the ICT. For these simulations, honest reporting was assumed.

Table 3

Forced Choice Technique Relative Reliability as Compared to the Direct Questioning Technique

	πs								
Sample Size	0.01	0.03	0.05	0.10	0.15	0.20	0.25	0.35	0.45
n = 150	0.03872	0.10672	0.16444	0.27624	0.3564	0.41592	0.4611	0.52236	0.55691
n = 500	0.03872	0.10672	0.16444	0.27624	0.3564	0.41592	0.4611	0.52236	0.55691
n = 1,500	0.03872	0.10672	0.16444	0.27624	0.3564	0.41592	0.4611	0.52236	0.55691

For each item list, relative reliability was compared to the DQT between the six models (i.e., three models with no purposeful correlations between non-sensitive questions and the three models with specific correlations between non-sensitive questions). These models are referred to as “Not-correlated” and “Correlated.” Results indicated that across list sizes, models that correlated at least one pair of non-sensitive questions proved more efficient (i.e., resulted in a higher relative reliability estimate as compared to DQT) compared to the models that did not correlate between non-sensitive questions. Because of this, the optimal model for each list size of the item count technique (ICT) was selected between the three correlated models. These results demonstrated what Glynn (2013) suggested that by purposefully correlating between non-sensitive questions, model variation decreased and statistical efficiency improved in the ICT.

Table 4 displays the relative reliability ratio for the three ICT correlated models by list size, sensitive prevalence rate, and sample size. The model with the greatest statistical efficiency was bolded in the table. In addition, Figures 2, 3, and 4 plot the relative reliability for each list size for the sensitive prevalence rates, $\pi_s = 0.05$, $\pi_s = 0.20$, and $\pi_s = 0.45$, when $n = 500$. These rates were selected to represent sensitive prevalent rates that are small, medium and large. Comparisons of the relative reliability among the three ICT correlated models indicated that the selection of the optimal model differed by the distribution of the prevalent rates of the non-sensitive questions (i.e., “equal,” “not equal but symmetric,” and “not equal and not symmetric”) and list size (3-item, 4-item, and 5-item). As was evident in both Table 4 and Figure 3, for the 4-item ICT model, results were consistent where the model with “equal” non-sensitive prevalence rates proved most efficient across all sensitive prevalence rates and sample size combinations. This was particularly evident in Figure 3, where relative reliability peaked when the distribution of the list of non-sensitive items is "equal" and correlated, indicating that in the even 4-item list correlation between similarly distributed sensitive questions reduced model variation compared to the models with correlated and unequal distributed sensitive questions (i.e., "not equal and symmetric," "not equal and not symmetric"). Recall that for the 4-item ICT list, when the distribution of the prevalent rate of the non-sensitive items was equal, two pairs of non-sensitive items could be negatively correlated whereas for the two unequal distributions, just one pair of non-sensitive questions could be correlated. This suggested that for even number item lists, correlating the maximum number of pairs of non-sensitive items improved efficiency compared to distributions of prevalent rates of non-sensitive items that were not equal where only a limited number of pairs of non-sensitive questions were negatively correlated. For the 3-item and 5-item list, statistical efficiency fluctuated between non-sensitive item lists that are distributed as "not equal but symmetric" and "not equal and not symmetric," indicating that unevenly distributed but correlated non-sensitive item lists reduced variation in item lists that are odd (i.e., 3-item and 5-item).

Table 4

Item Count Technique Relative Reliability, Selection of the Optimal Model

π_s	Item List Size	Non-Sensitive Distribution	Relative Reliability n = 150	Relative Reliability n = 500	Relative Reliability n = 1,500
0.01	3-Item (Cor)	Equal	0.0057	0.0056	0.0056
		Not Equal but Symmetric	0.0060	**0.0062**	**0.0062**
		Not Equal and Not Symmetric	**0.0062**	0.0058	0.0059
	4-Item (Cor)	Equal	**0.0050**	**0.0059**	**0.0057**
		Not Equal but Symmetric	0.0040	0.0039	0.0042
		Not Equal and Not Symmetric	0.0042	0.0041	0.0039
	5-Item (Cor)	Equal	0.0037	0.0039	0.0036
		Not Equal but Symmetric	**0.0044**	**0.0047**	0.0044
		Not Equal and Not Symmetric	0.0040	0.0046	**0.0046**
0.03	3-Item (Cor)	Equal	0.0161	0.0164	0.0146
		Not Equal but Symmetric	0.0172	0.0170	0.0186
		Not Equal and Not Symmetric	**0.0173**	**0.0173**	**0.0169**
	4-Item (Cor)	Equal	**0.0154**	**0.0158**	**0.0159**
		Not Equal but Symmetric	0.0121	0.0121	0.0118
		Not Equal and Not Symmetric	0.0125	0.0118	0.0122
	5-Item (Cor)	Equal	0.0105	0.0109	0.0106
		Not Equal but Symmetric	0.0121	0.0117	0.0127
		Not Equal and Not Symmetric	**0.0125**	**0.0128**	**0.0133**
0.05	3-Item (Cor)	Equal	0.0252	0.0249	0.0256
		Not Equal but Symmetric	0.0258	0.0255	**0.0285**
		Not Equal and Not Symmetric	**0.0281**	**0.0280**	0.0250
	4-Item (Cor)	Equal	**0.0229**	**0.0259**	**0.0272**
		Not Equal but Symmetric	0.0186	0.0206	0.0184
		Not Equal and Not Symmetric	0.0188	0.0187	0.0186
	5-Item (Cor)	Equal	0.0180	**0.0175**	0.0168
		Not Equal but Symmetric	**0.0208**	0.0191	**0.0210**
		Not Equal and Not Symmetric	0.0202	**0.0203**	0.0210
0.10	3-Item (Cor)	Equal	0.0438	0.0471	0.0450
		Not Equal but Symmetric	0.0497	0.0490	**0.0522**
		Not Equal and Not Symmetric	**0.0502**	**0.0492**	0.0506
	4-Item (Cor)	Equal	**0.0488**	**0.0450**	**0.0428**
		Not Equal but Symmetric	0.0351	0.0347	0.0335
		Not Equal and Not Symmetric	0.0329	0.0328	0.0340
	5-Item (Cor)	Equal	0.0319	0.0319	0.0304
		Not Equal but Symmetric	**0.0393**	**0.0381**	**0.0362**
		Not Equal and Not Symmetric	0.0383	0.0370	0.0361
0.15	3-Item (Cor)	Equal	0.0654	0.0637	0.0615
		Not Equal but Symmetric	**0.0699**	0.0650	0.0600
		Not Equal and Not Symmetric	0.0695	**0.0672**	**0.0713**
	4-Item (Cor)	Equal	**0.0614**	**0.0633**	**0.0602**
		Not Equal but Symmetric	0.0460	0.0494	0.0467
		Not Equal and Not Symmetric	0.0519	0.0494	0.0477
	5-Item (Cor)	Equal	0.0408	0.0422	0.0420
		Not Equal but Symmetric	**0.0563**	0.0480	**0.0552**
		Not Equal and Not Symmetric	0.0517	**0.0494**	0.0478

Table 4 (continued)

π_s	Item List Size	Non-Sensitive Distribution	Relative Reliability *n* = 150	Relative Reliability *n* = 500	Relative Reliability *n* = 1,500
0.20	3-Item (Cor)	Equal	0.0764	0.0747	0.0710
		Not Equal but Symmetric	**0.0869**	0.0806	**0.0895**
		Not Equal and Not Symmetric	0.0844	**0.0893**	0.0871
	4-Item (Cor)	Equal	**0.0712**	**0.0720**	**0.0749**
		Not Equal but Symmetric	0.0536	0.0579	0.0615
		Not Equal and Not Symmetric	0.0548	0.0557	0.0570
	5-Item (Cor)	Equal	0.0542	0.0556	0.0544
		Not Equal but Symmetric	0.0561	**0.0624**	**0.0627**
		Not Equal and Not Symmetric	**0.0628**	0.0619	0.0608
0.25	3-Item (Cor)	Equal	0.0830	0.0912	0.0920
		Not Equal but Symmetric	**0.0925**	0.0958	0.0932
		Not Equal and Not Symmetric	0.0898	**0.0960**	**0.0940**
	4-Item (Cor)	Equal	**0.0845**	**0.0810**	**0.0858**
		Not Equal but Symmetric	0.0678	0.0617	0.0660
		Not Equal and Not Symmetric	0.0645	0.0655	0.0665
	5-Item (Cor)	Equal	0.0578	0.0589	0.0673
		Not Equal but Symmetric	**0.0737**	0.0716	0.0714
		Not Equal and Not Symmetric	0.0696	**0.0756**	**0.0737**
0.35	3-Item (Cor)	Equal	0.1039	0.1041	0.1040
		Not Equal but Symmetric	**0.1145**	0.1078	0.1039
		Not Equal and Not Symmetric	0.1073	**0.1101**	**0.1095**
	4-Item (Cor)	Equal	**0.0964**	**0.1052**	**0.1071**
		Not Equal but Symmetric	0.0805	0.0767	0.0808
		Not Equal and Not Symmetric	0.0737	0.0752	0.0752
	5-Item (Cor)	Equal	0.0702	0.0736	0.0758
		Not Equal but Symmetric	0.0817	**0.0832**	**0.0811**
		Not Equal and Not Symmetric	**0.0818**	0.0815	0.0799
0.45	3-Item (Cor)	Equal	0.1059	0.1068	0.1086
		Not Equal but Symmetric	**0.1131**	**0.1241**	**0.1228**
		Not Equal and Not Symmetric	0.1121	0.1170	0.1182
	4-Item (Cor)	Equal	**0.1143**	**0.1063**	**0.1181**
		Not Equal but Symmetric	0.0873	0.0846	0.0941
		Not Equal and Not Symmetric	0.0877	0.0812	0.0850
	5-Item (Cor)	Equal	0.0762	0.0818	0.0729
		Not Equal but Symmetric	**0.0914**	0.0908	**0.0900**
		Not Equal and Not Symmetric	0.0840	**0.0963**	0.0869

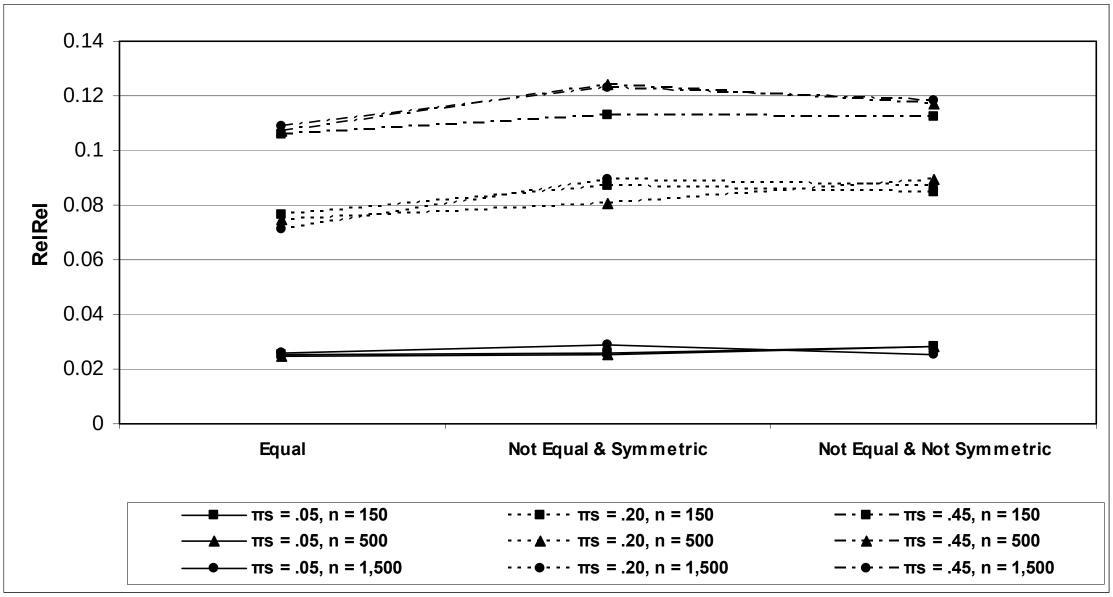


Figure 2. Item Count Technique Item List 3 (Correlated), Relative Reliability by Sensitive Prevalent Rate, and Distribution of Non-sensitive Prevalent Rates.

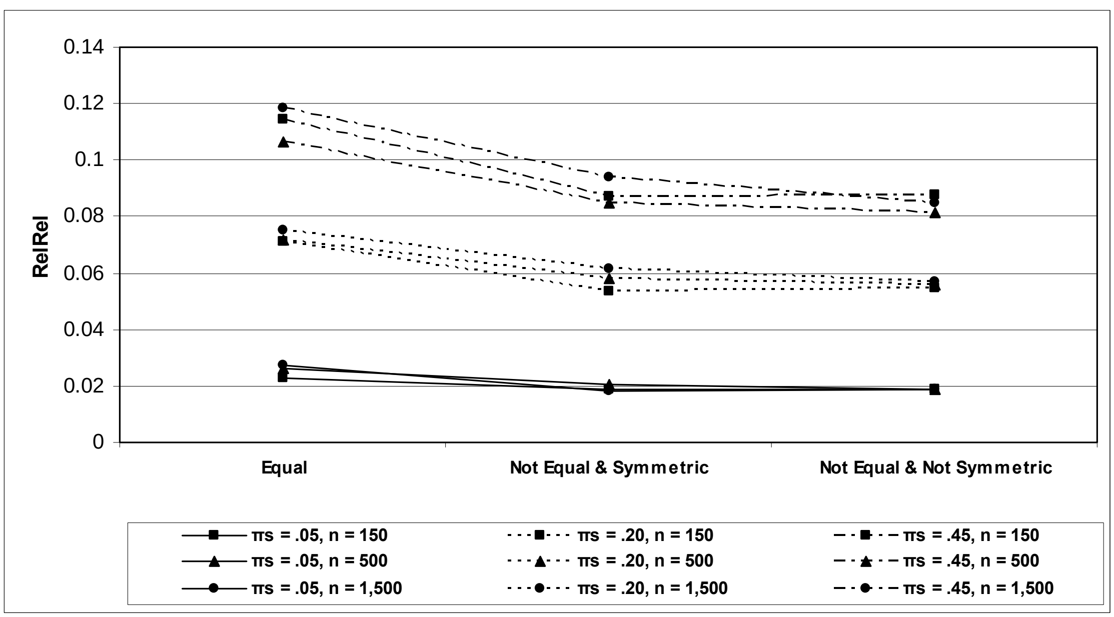


Figure 3. Item Count Technique Item List 4 (Correlated), Relative Reliability by Sensitive Prevalent Rate, and Distribution of Non-sensitive Prevalent Rates.

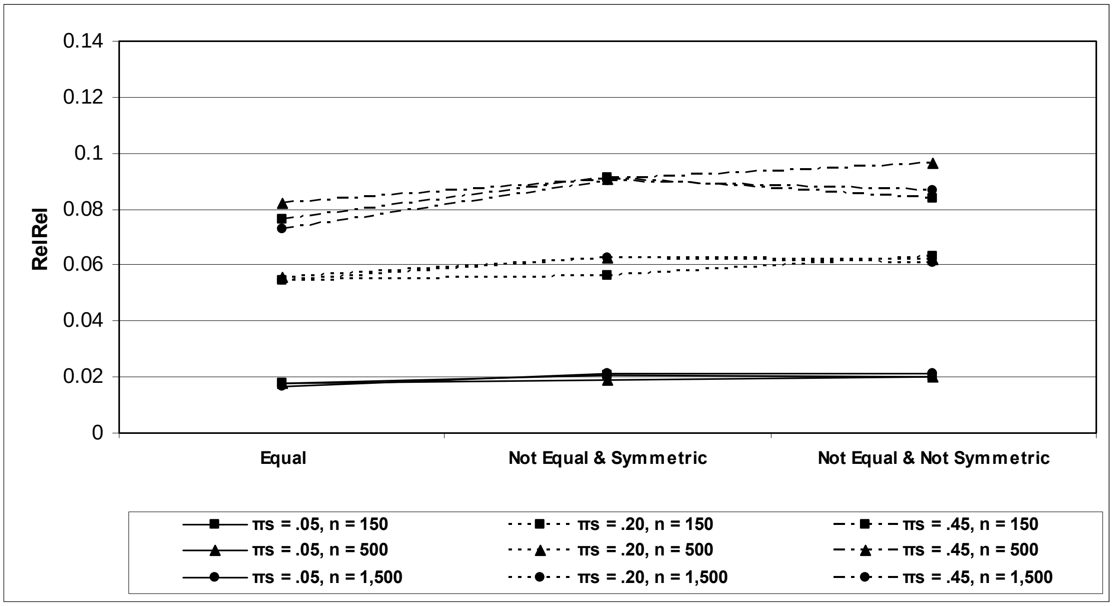


Figure 4. Item Count Technique Item List 5 (Correlated), Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates.

Because of the fluctuations of the relative reliability, further analyses were made in order to select the optimal model for both the 3-item and 5-item ICT model. The additional analyses examined the relative reliability *between* the models with non-sensitive item lists distributed as "equal and symmetric" and "not equal and not symmetric" since, as is shown in Table 4 and demonstrated in Figures 2 and 4, the optimal model for these list sizes oscillated between these two non-sensitive item distributions. For this analysis, relative reliability was measured in terms of the model with the maximum variation against the model with the minimum variation. So for instance, if the model with the "not equal and symmetric distribution" of non-sensitive items resulted in a larger variance compared with the "not equal and not symmetric distribution." the former was the numerator and the latter was the denominator. As a result, the measure was in terms of the additional percentage of sampled participants--

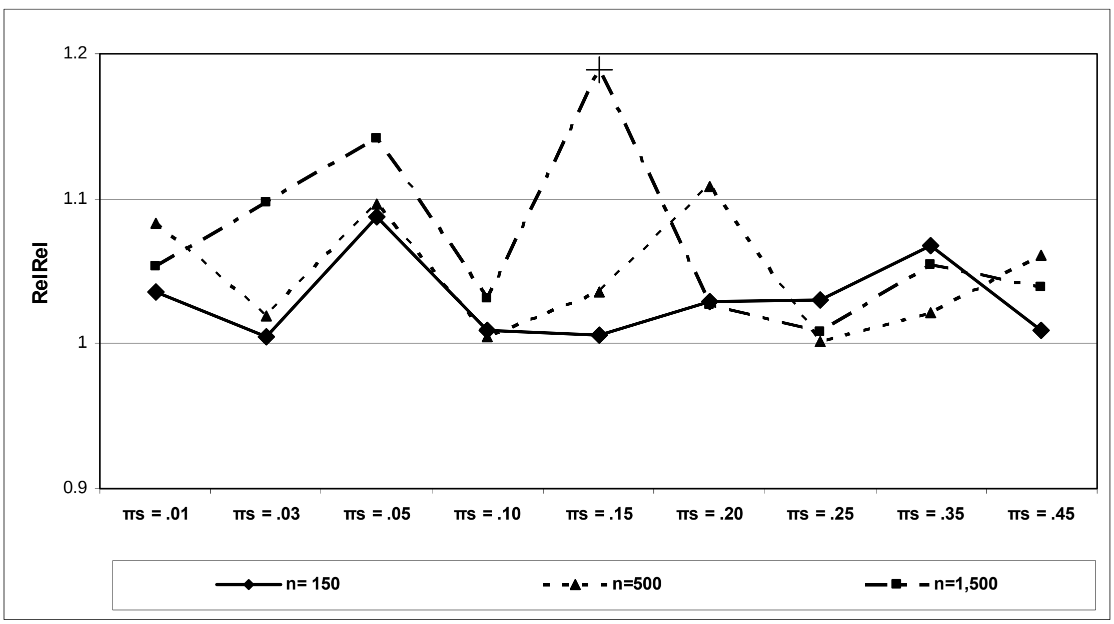
from the model with the higher variation--necessary to obtain a variance similar to the model with less variability. If the relative reliability was close to unity, either model could be termed as optimal. Table 5 displays the descriptive statistics of the relative reliability analysis between these two model types by item list; and Figures 5 and 6 display the plots of the relative reliability ratios by each sample size and item list, where the maximum relative reliability ratio is indicated in the plot by a "+". As is seen in the table, the relative reliability between the model types for the 3-item lists ranged between 1.00 and 1.19 with a mean and median of 1.05 and 1.04 and the relative reliability for the 5-item list ranged similarly between 1.00 and 1.16 where the mean and median were 1.05 and 1.03, respectively. For the ICT 3-item model and ICT 5-item model, the maximum relative reliability was 1.19 (not equal and not symmetric, n = 1,500, $\pi_s = 0.15$) and 1.16 ("not equal and symmetric," n = 1,500, $\pi_s = 0.15$). As a result, the optimal model for the 3-item and 5-item ICT, was selected based on the maximum relative reliability ratio, where the model with the smaller variability was selected. For the ICT 3-item model, the optimal mode was the correlated model with sensitive items distributed as "not equal and not symmetric" (max rel rel = 1.19) and for the ICT 5-item model, the optimal model was the correlated model with sensitive items distributed as "not equal and symmetric" (max rel rel = 1.16).

Table 5

Relative Reliability Study, Item Count Technique 3-Item and 5-Item Lists Comparing Models with "Not Equal and Symmetric" and "Not Equal and Not symmetric" Sensitive Prevalent Item Lists

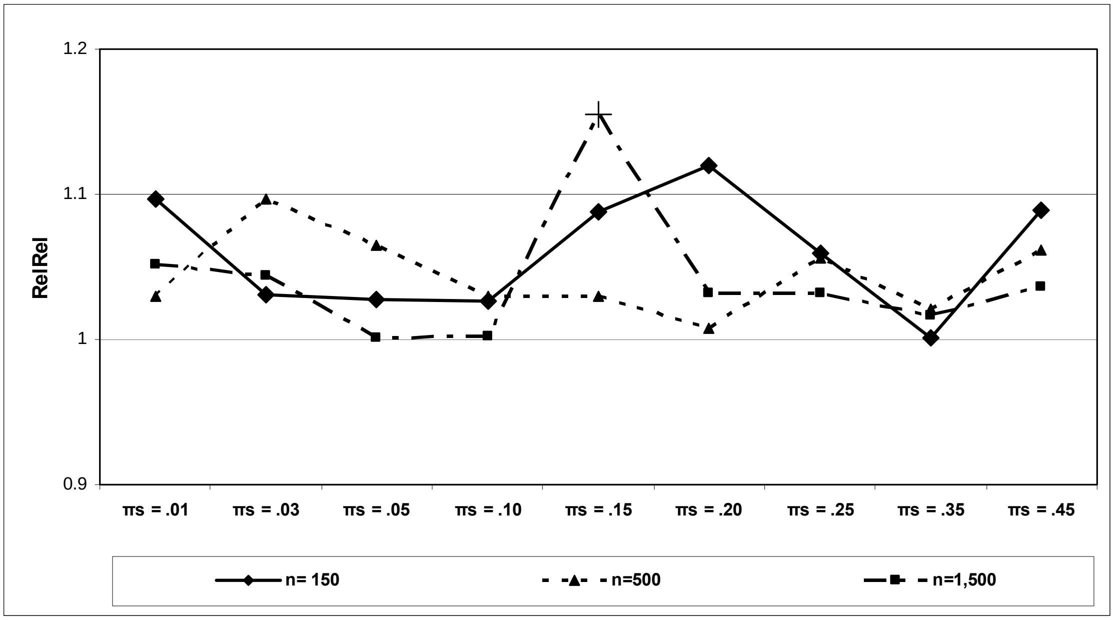
Item Size	Descriptive Statistics				
	Mean	Stddev	Median	Minimum	Maximum*
3-Item	1.0500	0.0464	1.0352	1.0013	1.1889
5-Item	1.0484	0.0382	1.0323	1.0007	1.1550

* Model with the maximum rel rel occurred for $\pi_s = 0.15$, $n = 1{,}500$. "not equal and not symmetric" (item list 3) and "not equal and symmetric" (item list 5)



+ Maximum rel rel occurred for $\pi_s = 0.15$, $n = 1{,}500$. "not equal and not symmetric"

Figure 5. Item Count Technique Item List 3 (Correlated), Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates (Not Equal and Symmetric vs. Not Equal and Not Symmetric)



+ Maximum rel rel occurred for $\pi_s = 0.15$, n = 1,500. "not equal and symmetric"

Figure 6. Item Count Technique Item List 5 (Correlated), Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive prevalent rates (Not Equal and Symmetric vs. Not Equal and Not Symmetric)

Double Item Count Technique (DICT)

Two sets of simulations were first attempted for the DICT in which non-sensitive questions were not purposely correlated (simulation 1) and selected non-sensitive questions correlated (simulation 2) as was done for the item count technique (ICT) described in the previous section. For both sets of simulations, *between* lists correlations were also simulated. Since at the time of this study's proposal, it was not certain if *between* lists correlations could be simulated, further research of the R-function "rmvbin" indicated that correlations *between* lists were possible. For the DICT, since both samples of participants received two lists, one with the sensitive question and one without, the defined correlation matrix per sample was larger than the correlation matrix per sample from the ICT simulations, in which only one sample received the sensitive question. For the simulated DICT lists, a *between* lists correlation of 0.85 implied that each non-sensitive item in list A was correlated with the corresponding non-sensitive item in list B. Thus, what distinguished the double item count technique (DICT) model from the item count technique (ICT) and single sample count technique (SSC) models in this study was that all non-sensitive items were correlated in pairs since there were two lists. As a result, for this study, *between* list correlations were made and in order to demonstrate the reduction of model variability due to the *between* list correlation, and yet be practical to real world applications, it was determined that a *between* list correlation of 0.85 (arbitrarily selected) would be adequate. When the two sets of simulations were attempted (i.e., simulations of *between* list correlations and simulations of *between* and *within* list correlations), the first set that considered correlations exclusively *between* lists, performed well. However, when correlations between non-sensitive items *within* lists were included, the multivariate normal approximation that would determine the threshold for the binary distribution resulted in a non-positive definite correlation matrix. As a result, simulations of the DICT that adjusted for both *within* and *between* item list correlations could not be done. Because of this, in order to determine optimal models for the DICT, an experiment was performed to determine which type of correlation--the *between* or *within*--reduced model variability more substantially in the DICT. This would then determine the correlation used in the set of simulations selecting the optimal DICT models.

Simulation Study of the None vs. Within vs. Between Correlation in the 5-item Double Item Count Technique

For this simulation study, the model with the highest expected variability was used. The 5-item DICT model for samples of 150 subjects was selected since it is expected to produce the highest amount of variability (Glynn, 2013; Tsuchiya et al., 2007). Simulations were run for each of the nine sensitive prevalent rates in order to examine the models variability more thoroughly.

A total of three sets of simulations were defined by type of correlation (i.e., *none*, *between list, within list*). The first set did not purposely correlate either *between* lists or *within* lists; whereas the second set correlated *between* lists at 0.85 and no *within* list correlations, and the third set correlated *within* lists at -0.50 as outlined in Chapter III for the item count technique (ICT) and no *between* list correlations. For each of these correlation types, a total of 1,000 simulations were run by sensitive prevalent rate in combination with non-sensitive prevalent rate distributions ("equal," "not equal and symmetric," "not equal and not symmetric"), resulting in a total of 81 simulations (27 simulations per correlation type). For each simulation, 1,000 double item count technique (DICT) 5-item lists per sample were simulated using the "rmvbin" function in R, resulting in 1,000 DICT estimates. Variances were than calculated. Using these variances, the relative reliability between correlation type models (*none*, *between*, and *within*) as defined by non-sensitive distribution type ("equal," "not equal and symmetric," and "not equal and not symmetric") were taken for each sensitive prevalent rate. So for instance, comparisons between the variance of the 5-item DICT estimating sensitive prevalent rate $\pi_s = 0.01$ where the non-sensitive item distribution was "equal" were compared between models with no correlation, *between* correlation and *within* correlation. For this set of simulations, truthful reporting was assumed. As was done in the previous set of simulations used to determine the optimal model for the 3-item and 5-item ICT, the relative reliability ratio measured the model with the maximum variation (numerator) versus the model with the minimum variation (denominator). So for instance, if the model with "no correlation" resulted in a larger variance compared with the model with "*between*" correlation, the former would be used as the numerator and the latter as the denominator. As a result, the measure would be in terms of the percent of sampled participants--from the model with the higher variation--necessary to obtain a variance similar to the model with less variability. If the relative reliability was not close to unity, the model with the smaller variability would be termed more efficient and used in the set of simulations to determine the optimal model for the double item count technique (DICT). Table 6 displays the results of the simulation study. As can be seen in the table, the relative reliability of the models indicates that DICT models with highly correlated *between* lists reduced model variability compared with DICT models *within* or *no* list correlation. This is apparent since the relative reliability across sensitive prevalent rates and non-sensitive item distribution type was substantially greater than unity. Because of this, for the simulations determining the optimal DICT model, no *within* list correlations were made. Instead only *between* list correlations were simulated and the optimal model was selected by the distribution of the non-sensitive prevalent rates (i.e., "equal," "not equal and symmetric," and "not equal and not symmetric").

Table 6

Double Item Count Technique Relative Reliability, Study of Within List, Between List, and No Correlation in the 5-Item Model

Dist Type	π_s	Correlation			Relative Reliability: None vs. Within	Relative Reliability: None vs. Between	Relative Reliability: Within vs. Between
		Variance Between	Variance Within	Variance None			
Equal	0.01	0.0024	0.0093	0.0137	1.4734	5.6804	3.8552
	0.03	0.0026	0.0090	0.0153	1.7027	5.8865	3.4571
	0.05	0.0026	0.0091	0.0148	1.6362	5.7024	3.4852
	0.10	0.0031	0.0093	0.0168	1.8063	5.4338	3.0082
	0.15	0.0032	0.0098	0.0157	1.6042	4.8507	3.0237
	0.20	0.0035	0.0097	0.0160	1.6560	4.6260	2.7935
	0.25	0.0039	0.0100	0.0160	1.6009	4.1280	2.5785
	0.35	0.0040	0.0105	0.0163	1.5464	4.0708	2.6324
	0.45	0.0038	0.0106	0.0151	1.4303	3.9412	2.7555
Not Equal & Symm	0.01	0.0021	0.0075	0.0139	1.8426	6.5532	3.5565
	0.03	0.0023	0.0073	0.0126	1.7161	5.5463	3.2319
	0.05	0.0024	0.0080	0.0131	1.6227	5.3869	3.3198
	0.10	0.0026	0.0080	0.0133	1.6606	5.1459	3.0988
	0.15	0.0027	0.0090	0.0137	1.5192	5.0179	3.3029
	0.20	0.0033	0.0080	0.0135	1.6752	4.1152	2.4565
	0.25	0.0030	0.0092	0.0146	1.5937	4.8602	3.0496
	0.35	0.0036	0.0090	0.0141	1.5667	3.8780	2.4753
	0.45	0.0038	0.0092	0.0138	1.5038	3.6286	2.4130
Not Equal & Not Symm	0.01	0.0023	0.0077	0.0137	1.7750	5.9153	3.3325
	0.03	0.0025	0.0079	0.0143	1.8089	5.6380	3.1169
	0.05	0.0024	0.0072	0.0140	1.9430	5.7377	2.9529
	0.10	0.0029	0.0080	0.0147	1.8495	5.0777	2.7454
	0.15	0.0032	0.0081	0.0139	1.7242	4.3868	2.5443
	0.20	0.0034	0.0097	0.0154	1.5913	4.4914	2.8225
	0.25	0.0033	0.0089	0.0166	1.8615	5.0228	2.6983
	0.35	0.0038	0.0088	0.0151	1.7205	3.9628	2.3033
	0.45	0.0038	0.0097	0.0140	1.4404	3.7033	2.5710

Selecting the Optimal Double Item Count Technique Model

Table 7 displays the relative reliability as compared with the DQT for the three double item count technique (DICT) *between* list correlated models by list size, sensitive prevalence rate and sample size. The model with the greatest efficiency is bolded in the table per sample size and sensitive prevalent rate combination. In addition, Figures 7, 8, and 9 plot the relative reliability for each list size for sensitive prevalent rates $\pi_s = 0.05$ (small), $\pi_s = 0.20$ (medium), and $\pi_s = 0.45$ (large). Comparisons of the relative reliability among the three DICT correlated models indicated no consistent results in the selection of the optimal model by the distribution of the prevalent rates of the non-sensitive questions (i.e., “equal,” “not equal but symmetric,” “not equal and not symmetric”) and list size (3-item, 4-item, and 5-item). As is evident in the table, efficiency for the 3-item and 4-item DICT fluctuated between each of the non-sensitive item distributions (i.e., “equal,” “not equal and symmetric,” and “not equal and not symmetric”). Figures 7 and 8 also indicated no clear differences between the efficiency of the non-sensitive distribution types since the plot of the relative reliability appear flat and do not peak at any one list type. This was evident across all sensitive prevalent rates and sample sizes. For the 5-item DICT, Table 7 indicated that efficient models fluctuated between the “not equal and symmetric” and “not equal and not symmetric” distribution of the non-sensitive items. The models with “equally” distributed non-sensitive questions were never selected as optimal. Figure 9, which plots the relative reliability for the 5-item DICT, demonstrated this since the plot peaked at either the “not equal and symmetric” or “not equal and not symmetric” models and valleyed at the “equal” models.

Table 7

Double Item Count Technique Relative Reliability, Selection of the Optimal Model

π_s	Item List Size	Non-Sensitive Distribution	*n* = 150	*n* = 500	*n* = 1,500
0.01	3- tem (Cor Bet)	Equal	0.0429	0.0447	0.0436
		Not Equal but Symmetric	0.0436	0.0441	0.0422
		Not Equal and Not Symmetric	**0.0455**	**0.0466**	**0.0464**
	4-Item (Cor Bet)	Equal	0.0330	0.0390	0.0354
		Not Equal but Symmetric	0.0321	0.0345	**0.0369**
		Not Equal and Not Symmetric	**0.0352**	**0.0347**	0.0357
	5-Item (Cor Bet)	Equal	0.0287	0.0277	0.0252
		Not Equal but Symmetric	**0.0315**	**0.0297**	**0.0324**
		Not Equal and Not Symmetric	0.0294	0.0285	0.0295
0.03	3-Item (Cor Bet)	Equal	0.1165	**0.1298**	0.1193
		Not Equal but Symmetric	0.1129	0.1099	0.1135
		Not Equal and Not Symmetric	**0.1178**	0.1240	**0.1245**
	4-Item (Cor Bet)	Equal	0.0941	0.0978	0.0873
		Not Equal but Symmetric	0.0873	**0.0979**	**0.0984**
		Not Equal and Not Symmetric	**0.0965**	0.0928	0.0979
	5-Item (Cor Bet)	Equal	0.0748	0.0727	0.0744
		Not Equal but Symmetric	**0.0880**	0.0863	**0.0859**
		Not Equal and Not Symmetric	0.0773	**0.0865**	0.0767
0.05	3-Item (Cor Bet)	Equal	0.1696	0.1808	**0.1933**
		Not Equal but Symmetric	0.1805	**0.1809**	0.1726
		Not Equal and Not Symmetric	**0.1910**	0.1799	0.1878
	4-Item (Cor Bet)	Equal	0.1398	0.1492	0.1310
		Not Equal but Symmetric	**0.1489**	0.1493	0.1350
		Not Equal and Not Symmetric	0.1421	**0.1503**	**0.1553**
	5-Item (Cor Bet)	Equal	0.1156	0.1184	0.1169
		Not Equal but Symmetric	0.1285	**0.1426**	**0.1406**
		Not Equal and Not Symmetric	**0.1351**	0.1261	0.1327
0.10	3-Item (Cor Bet)	Equal	0.3138	**0.3057**	**0.3128**
		Not Equal but Symmetric	0.2752	0.2980	0.2761
		Not Equal and Not Symmetric	**0.2758**	0.2756	0.2962
	4-Item (Cor Bet)	Equal	**0.2424**	0.2332	**0.2523**
		Not Equal but Symmetric	0.2424	0.2467	0.2405
		Not Equal and Not Symmetric	0.2368	**0.2698**	0.2388
	5-Item (Cor Bet)	Equal	0.2181	0.2118	0.2125
		Not Equal but Symmetric	**0.2316**	**0.2213**	0.2071
		Not Equal and Not Symmetric	0.2015	0.1972	**0.2157**

Table 7 (continued)

π_s	Item List Size	Non-Sensitive Distribution	*n* = 150	*n* = 500	*n* = 1,500
0.15	3-Item (Cor Bet)	Equal	**0.3958**	**0.4002**	0.3623
		Not Equal but Symmetric	0.3915	0.3680	0.3625
		Not Equal and Not Symmetric	0.3908	0.3868	**0.3734**
	4-Item (Cor Bet)	Equal	0.2928	0.3142	0.3138
		Not Equal but Symmetric	0.3081	0.3117	**0.3232**
		Not Equal and Not Symmetric	**0.3406**	**0.3217**	0.3220
	5-Item (Cor Bet)	Equal	0.2609	0.2762	0.2791
		Not Equal but Symmetric	0.2796	**0.2994**	**0.3030**
		Not Equal and Not Symmetric	**0.2813**	0.2824	0.2760
0.20	3-Item (Cor Bet)	Equal	0.4293	**0.4506**	0.3949
		Not Equal but Symmetric	0.4294	0.4179	**0.4438**
		Not Equal and Not Symmetric	**0.4564**	0.4478	0.4348
	4-Item (Cor Bet)	Equal	0.3773	0.3589	0.3537
		Not Equal but Symmetric	0.3512	**0.3710**	0.3741
		Not Equal and Not Symmetric	**0.3793**	0.3410	**0.3956**
	5-Item (Cor Bet)	Equal	0.2922	0.3000	0.2989
		Not Equal but Symmetric	**0.3518**	**0.3716**	0.3257
		Not Equal and Not Symmetric	0.3336	0.3229	**0.3594**
0.25	3-Item (Cor Bet)	Equal	0.4712	**0.4566**	0.4660
		Not Equal but Symmetric	0.4420	0.4780	**0.4793**
		Not Equal and Not Symmetric	**0.4904**	0.5080	0.4454
	4-Item (Cor Bet)	Equal	0.3632	0.3975	0.3886
		Not Equal but Symmetric	0.4000	0.4193	**0.4058**
		Not Equal and Not Symmetric	**0.4105**	**0.4303**	0.4034
	5-Item (Cor Bet)	Equal	0.3287	0.3280	0.3463
		Not Equal but Symmetric	**0.3953**	**0.4064**	0.3739
		Not Equal and Not Symmetric	0.3354	0.3707	**0.3890**
0.35	3-Item (Cor Bet)	Equal	**0.5507**	**0.5376**	0.4928
		Not Equal but Symmetric	0.5367	0.4940	0.4924
		Not Equal and Not Symmetric	0.5153	0.4951	**0.5051**
	4-Item (Cor Bet)	Equal	**0.4748**	0.4501	0.4620
		Not Equal but Symmetric	0.4414	0.4291	0.4065
		Not Equal and Not Symmetric	0.4339	**0.4592**	**0.4497**
	5-Item (Cor Bet)	Equal	0.3880	0.3797	0.4005
		Not Equal but Symmetric	**0.4398**	**0.4601**	**0.4548**
		Not Equal and Not Symmetric	0.4219	0.4181	0.3991
0.45	3-Item (Cor Bet)	Equal	0.4929	0.5351	**0.5268**
		Not Equal but Symmetric	**0.5420**	**0.5379**	0.5236
		Not Equal and Not Symmetric	0.4932	0.5181	0.5243

Table 7 (continued)

π_s	Item List Size	Non-Sensitive Distribution	*n* = 150	*n* = 500	*n* = 1,500
	4-Item (Cor Bet)	Equal	0.4429	0.4529	0.4464
		Not Equal but Symmetric	0.4453	0.4729	0.4825
		Not Equal and Not Symmetric	**0.4636**	**0.5227**	**0.5055**
	5-Item (Cor Bet)	Equal	0.3791	0.4090	0.3941
		Not Equal but Symmetric	0.4018	**0.4396**	0.4603
		Not Equal and Not Symmetric	**0.4537**	0.4333	**0.4746**

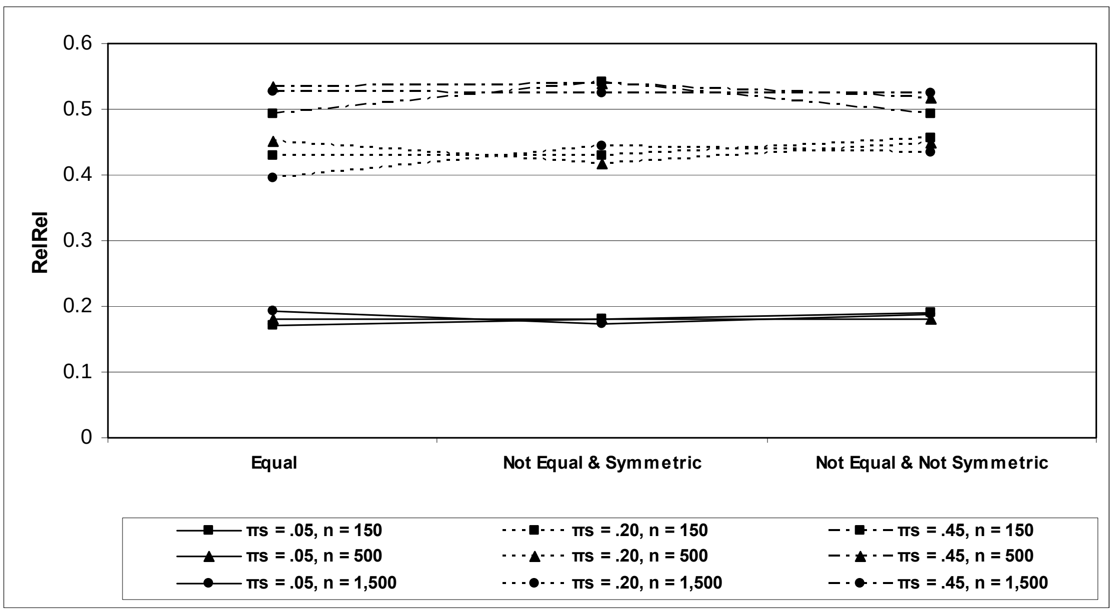


Figure 7. Double Item Count Technique Item List 3, Relative Reliability by Sensitive Prevalent Rate, and Distribution of Non-sensitive Prevalent Rates.

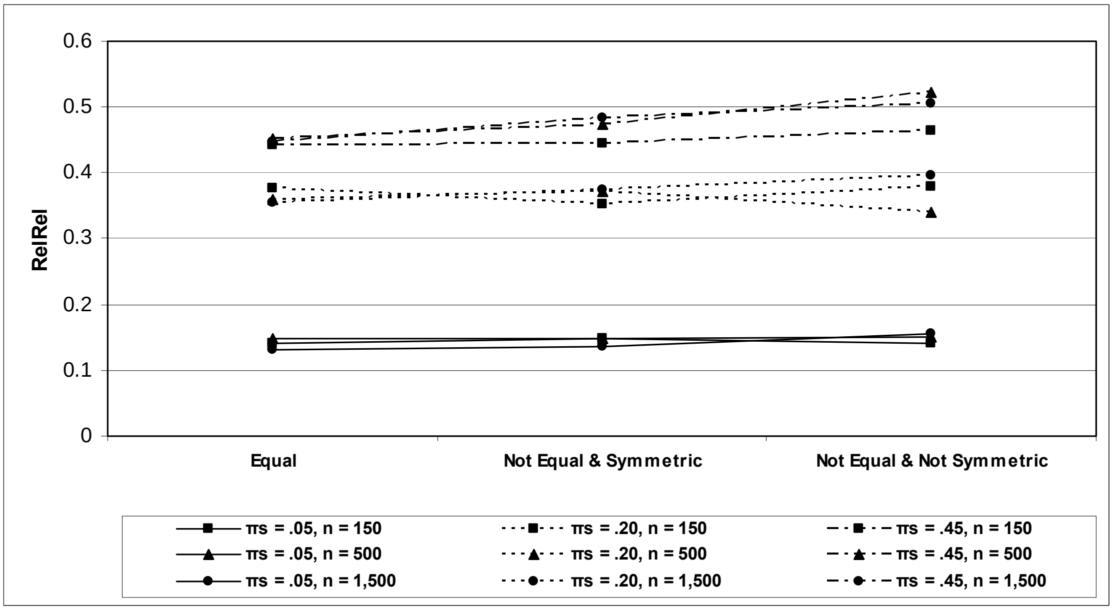


Figure 8. Double Item Count Technique Item List 4, Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates.

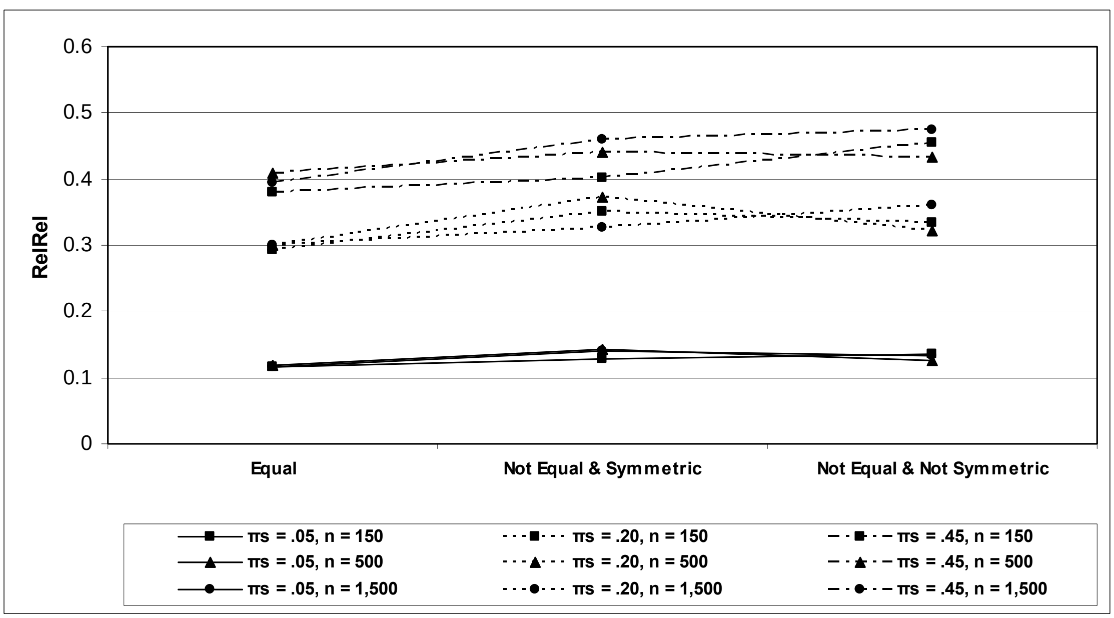


Figure 9. Double Item Count Technique Item List 5, Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates.

As a result of this, a relative reliability analysis similar to the one performed for the 3-item and 5-item item count technique (ICT) was run to determine the optimal model. For this analysis, since all three non-sensitive distribution types were selected at some point for the 3-item and 4-Item double item count technique (DICT), all model types were included in the analysis. For the 5-item DICT, since models for which the non-sensitive questions were distributed "equally" were never selected, the relative reliability was examined between models where non-sensitive questions were distributed as "not equal and symmetric" and "not equal and not symmetric." Relative reliability was measured in terms of the ratio of the model with the higher variability compared to the model with the smaller variability. In order to encompass all model types, for the 3-item and 4-item DICT, the model resulting in the maximum variance was compared to the model resulting in the minimum variance. For the 5-item DICT, the comparison was similar, where the comparison occurred between the models with non-sensitive prevalent rates distributed as "not equal and symmetric" to the models with non-sensitive prevalent rates distributed as "not equal and not symmetric." As a result, the measures are in terms of the additional percentage of sampled participants--from the model with the higher variation--necessary to obtain a variance similar to the model with less variability. If the relative reliability was close to unity, either model could be termed as optimal. Table 8 displays the descriptive statistics of the relative reliability analysis for the DICT by item list size; and Figures 10, 11, and 12 display the plots of the relative reliability ratios by each sample size and item list. In each plot, the maximum relative reliability is indicated by a "+". As is seen in the table, the relative reliability between the model types for the 3-item lists ranged between 1.01 and 1.18 with a mean and median of 1.08 and 1.09;

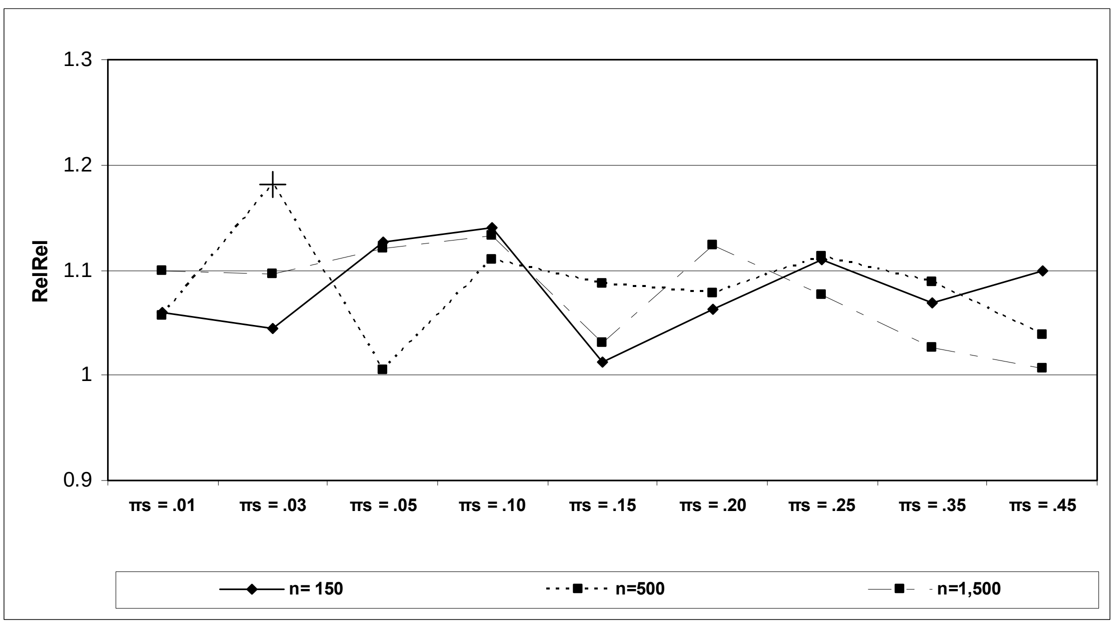
whereas the relative reliability for the 4-item list ranged similarly between 1.01 and 1.19 with mean and median 1.09. For the 5-item double item count technique (DICT), the relative reliability also ranged similarly between 1.00 and 1.18 where the mean and median were 1.08 and 1.06, respectively. For the DICT 3-item, 4-item, and 5-item models, the maximum relative reliability was 1.18 (comparing "not equal and symmetric" to "equal" to, n = 500, $\pi_s = 0.03$), 1.19 (comparing "equal" and "not equal and not symmetric," n = 1,500, $\pi_s = 0.05$), and 1.18 (comparing "not equal and not symmetric" to "not equal and symmetric," n = 150, $\pi_s = 0.25$) indicating that, at maximum, between 18 and 19 percent additional subjects were necessary in order to produce equivalent models. As a result, the optimal model for the 3-item, 4-item, and 5-item DICT, was based on the maximum relative reliability ratio, where the model with the smaller variability was selected. For the DICT 3-item model, the optimal model was the *between* list correlated model with sensitive items distributed as "equal" (max rel rel = 1.18) and for the DICT 4-item model, the optimal model was the *between* list correlated model with sensitive items distributed as "not equal and not symmetric" (max rel rel = 1.19). The optimal model selected for the DICT 5-item was the model with the *between* list correlation with sensitive items distributed as "not equal and symmetric" (max rel rel = 1.18).

Table 8

Relative Reliability Study, Double Item Count Technique 3-Item, 4-Item, and 5-Item Lists Comparing Models with "Equal," "Not Equal and Symmetric," and "Not Equal and Not Symmetric" Sensitive Prevalent Item Lists

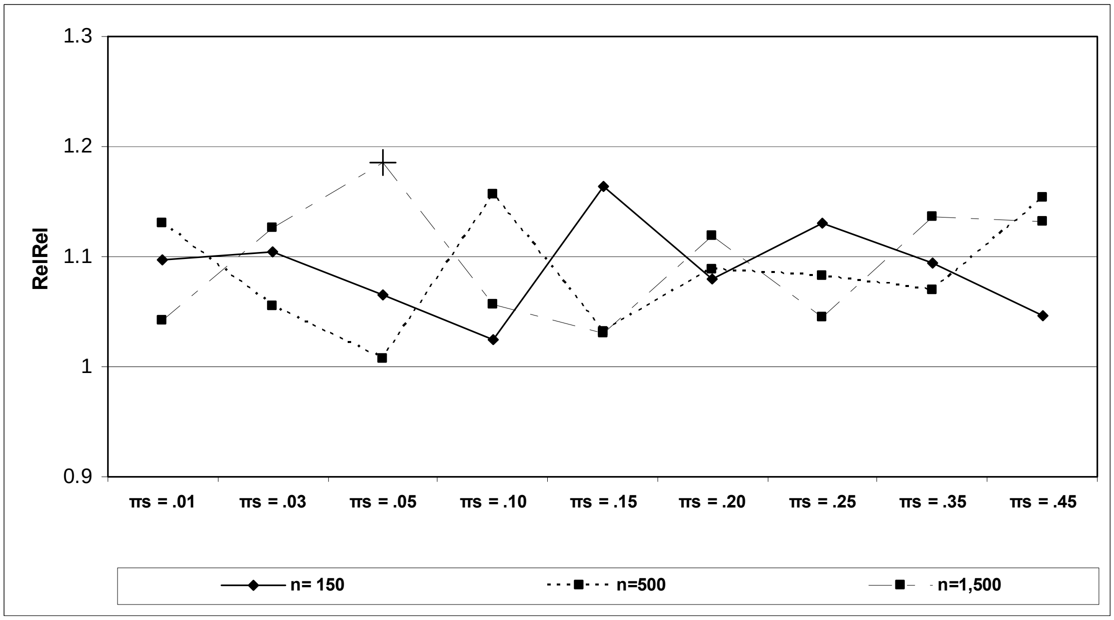
Item Size	Descriptive Statistics				
	Mean	Stddev	Median	Minimum	Maximum
3-Item	1.0812	0.0445	1.0875	1.0057	1.1813
4-Item	1.0909	0.0483	1.0878	1.0078	1.1850
5-Item	1.0764	0.0479	1.0621	1.0019	1.1786

*Model with the maximum rel rel occurred for $\pi_s = 0.03$, $n = 500$. "equal" (item list 3), $\pi_s = 0.05$, $n = 1{,}500$. "not equal and not symmetric" (item list 4), $\pi_s = 0.25$, $n = 150$. "not equal and symmetric" (item list 5)



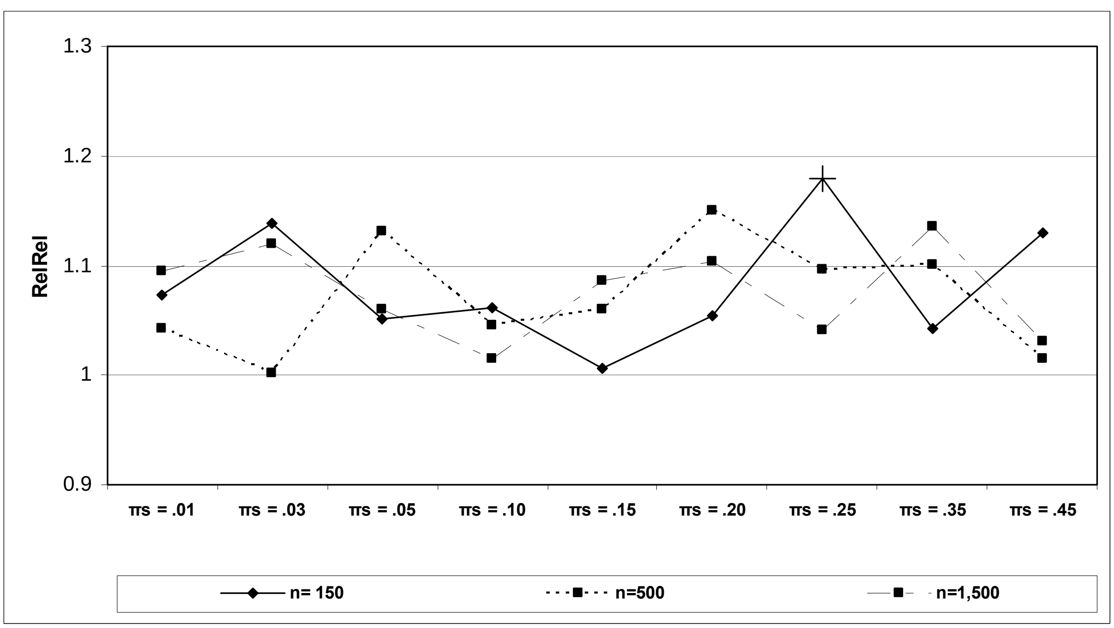
+ Maximum rel rel occurred for $\pi_s = 0.03$, $n = 500$. "equal"

Figure 10. Double Item Count Technique Item List 3, Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates (Equal vs. Not Equal, and Symmetric vs. Not Equal and Not Symmetric)



+ Maximum rel rel occurred for $\pi_s = 0.05$, $n = 1{,}500$. "not equal and not symmetric"

Figure 11. Double Item Count Technique Item List 4, Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates (Equal vs. Not Equal and Symmetric vs. Not Equal and Not Symmetric)



+ Maximum rel rel occurred for $\pi_s = 0.25$, $n = 150$. "not equal and symmetric"

Figure 12. Double Item Count Technique Item List 5, Relative Reliability by Sensitive Prevalent Rate and Distribution of Non-sensitive Prevalent Rates (Not Equal and Symmetric vs. Not Equal and Not symmetric)

Single Sample Count Technique (SSC)

For the single sample count technique (SSC), two sets of simulations were run. In the first set, simulations of non-sensitive questions were not purposely correlated and in the second set, specific non-sensitive questions were correlated at -0.50 as outlined in Chapter III. Simulations were then performed for each prevalent rate (with the exception of π_s = 0.01) - sample size combination. Since the distribution of prevalent rates of the non-sensitive items were equal and set at 0.50 for all non-sensitive questions, distribution of the sensitive questions was not a factor in these simulations. Thus, a total of 48 simulations per list size were run. For each simulation, 1,000 item lists were created using the R function, “rmvbin,” resulting in a total of 1,000 expected value estimates for which the variance was calculated and used as an estimate of the variability in the SSC. For these simulations, honest reporting was assumed.

For each item list, relative reliability was compared to the DQT between the two models (i.e., the model with no purposeful correlations between non-sensitive questions and the model with specific correlations between non-sensitive questions). Table 9 displays the results of the relative reliability analysis. Optimal models by list size and sensitive prevalent rate are bolded in the table. Results indicated that across list sizes and sensitive prevalent rates, models that correlated at least one pair of non-sensitive questions proved more efficient (i.e., resulted in a higher relative reliability estimate as compared to DQT) compared to the models that did not correlate between non-sensitive questions. Because of this, the optimal model for each list size of the SSC was the model that correlated between at least one pair of sensitive questions. Like the ICT and DICT, these results demonstrated what Glynn (2013) suggested that by purposefully correlating between non-sensitive questions, model variation decreased and statistical efficiency improved.

Table 9

Single Sample Count Relative Reliability, Selection of the Optimal Model

π_s	List Size	Correlated			Not Correlated		
		n = 150	*n* = 500	*n* = 1,500	*n* = 150	*n* = 500	*n* = 1,500
0.03	3-Item	**0.054563**	**0.057512**	**0.057261**	0.038832	0.039217	0.035902
	4-Item	**0.048644**	**0.057113**	**0.0594859**	0.028011	0.029640	0.027327
	5-Item	**0.036840**	**0.035354**	**0.036288**	0.022017	0.022322	0.021286
0.05	3-Item	**0.078435**	**0.084802**	**0.0860270**	0.059609	0.059331	0.057416
	4-Item	**0.087302**	**0.084969**	**0.095583**	0.047117	0.046825	0.04518
	5-Item	**0.057595**	**0.060088**	**0.0569652**	0.036235	0.039181	0.036715
0.10	3-Item	**0.153976**	**0.148668**	**0.1414304**	0.117482	0.101291	0.106261
	4-Item	**0.157227**	**0.160248**	**0.1624203**	0.082637	0.082438	0.084776
	5-Item	**0.104396**	**0.109390**	**0.1120014**	0.064650	0.068787	0.069452
0.15	3-Item	**0.209415**	**0.209558**	**0.2106366**	0.144481	0.145469	0.140555
	4-Item	**0.203272**	**0.201227**	**0.2182143**	0.120184	0.107018	0.107828
	5-Item	**0.146633**	**0.144657**	**0.160424**	0.083426	0.087641	0.088985
0.20	3-Item	**0.238186**	**0.273521**	**0.2307898**	0.191087	0.165385	0.187770
	4-Item	**0.22999**	**0.237275**	**0.2390333**	0.126577	0.138758	0.134126
	5-Item	**0.176778**	**0.164041**	**0.1799423**	0.104794	0.106479	0.118225
0.25	3-Item	**0.282802**	**0.278962**	**0.2707125**	0.197648	0.190178	0.202867
	4-Item	**0.268962**	**0.270955**	**0.2549287**	0.154700	0.167569	0.167614
	5-Item	**0.207385**	**0.203666**	**0.1956293**	0.129096	0.125561	0.127687
0.35	3-Item	**0.319831**	**0.325236**	**0.3159866**	0.256686	0.234091	0.232354
	4-Item	**0.328424**	**0.318911**	**0.3204324**	0.194757	0.192235	0.187512
	5-Item	**0.235264**	**0.236536**	**0.2198946**	0.159811	0.147532	0.168111
0.45	3-Item	**0.334170**	**0.327620**	**0.3101452**	0.235714	0.256463	0.241387
	4-Item	**0.338029**	**0.325328**	**0.3385486**	0.194628	0.200729	0.190010
	5-Item	**0.247533**	**0.249735**	**0.2413803**	0.175380	0.164462	0.171289

The Study of Non-Compliance

Unrelated Question Technique (Optimal) vs. Non-random Response Techniques (Optimal)

Table 10 displays the results of the non-compliance study, comparing the optimal UQT technique to each of the NRR optimal techniques. Figures 13, 14, and 15 display the ratio of MSE by the proportion of truthful reporting (0.90, 0.80, 0.70, and 0.60) and sample size (150, 500, and 1,500) for sensitive prevalent rates $\pi_s = 0.05$ (small), $\pi_s = 0.20$ (medium), and $\pi_s = 0.45$ (large). In each figure, the plot of the ratio of MSE by NRR technique is coded by line type, where plots of all item count technique (ICT) models are solid-broken, double item count technique (DICT) models are solid and single sample count technique (SSC) models are broken. Item list size are represented by plot type where the 3-item list is represented by triangles, the 4-item list represented by circles and the 5-item list is represented by squares. For these analyses, the study assumed that a ratio of MSE close to unity implied the models were similar in their efficiency.

As is evident in the table and figures, when π_s was small (i.e., 0.01, 0.03, and 0.05) across all sample sizes, the UQT optimal model was generally more efficient than all NRR models.

When π_s was moderate (i.e., 0.10, 0.15, and 0.20), as π_s increases the DICT optimal models approached the efficiency of the UQT for mid sample sizes ($n = 500$) and large sample sizes ($n = 1,500$) in the presence of high non-compliance (i.e., percent of truthful reporting ≤ 0.70). For the optimal SSC, in the presence of high non-compliance (i.e., percent of truthful reporting ≤ 0.70), as π_s increased for larger sample sizes (i.e., 1,500), the model was nearly as efficient as the UQT optimal.

Table 10

Study of Non-compliance: Unrelated Question Technique Optimal vs. Non-random Response Optimal Models

π_s	Non-random Response Model	*n* = 150				*n* = 500				*n* = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.01	ICT 3-Item	0.022	0.022	0.022	0.019	0.021	0.021	0.022	0.022	0.021	0.023	0.030	0.034
	ICT 4-Item	0.021	0.019	0.019	0.018	0.019	0.019	0.020	0.022	0.020	0.022	0.027	0.031
	ICT 5-Item	0.014	0.015	0.014	0.015	0.017	0.016	0.017	0.016	0.015	0.017	0.021	0.024
	DICT 3-Item	0.150	0.152	0.145	0.148	0.146	0.148	0.153	0.167	0.160	0.163	0.189	0.199
	DICT 4-Item	0.126	0.120	0.116	0.127	0.120	0.125	0.128	0.138	0.132	0.139	0.148	0.175
	DICT 5-Item	0.099	0.104	0.098	0.100	0.102	0.110	0.122	0.111	0.106	0.120	0.138	0.179
	SSC 3-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 4-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 5-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.03	ICT 3-Item	0.035	0.035	0.035	0.038	0.037	0.044	0.054	0.070	0.044	0.063	0.090	0.137
	ICT 4-Item	0.031	0.032	0.034	0.039	0.035	0.039	0.047	0.061	0.038	0.056	0.093	0.123
	ICT 5-Item	0.026	0.024	0.026	0.031	0.028	0.032	0.041	0.050	0.030	0.044	0.074	0.110
	DICT 3-Item	0.244	0.241	0.244	0.250	0.247	0.287	0.345	0.418	0.278	0.344	0.484	0.647
	DICT 4-Item	0.186	0.194	0.212	0.224	0.195	0.231	0.268	0.310	0.236	0.325	0.404	0.487
	DICT 5-Item	0.178	0.169	0.185	0.200	0.187	0.201	0.273	0.293	0.203	0.283	0.351	0.463
	SSC 3-Item	0.116	0.122	0.119	0.130	0.102	0.145	0.170	0.198	0.138	0.195	0.281	0.353
	SSC 4-Item	0.117	0.116	0.113	0.133	0.124	0.139	0.167	0.206	0.132	0.187	0.282	0.362
	SSC 5-Item	0.074	0.074	0.081	0.081	0.081	0.093	0.105	0.137	0.089	0.133	0.201	0.274

Table 10 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.05	ICT 3-Item	0.045	0.053	0.058	0.072	0.052	0.075	0.097	0.139	0.068	0.127	0.198	0.284
	ICT 4-Item	0.048	0.047	0.059	0.069	0.046	0.072	0.087	0.125	0.060	0.116	0.192	0.295
	ICT 5-Item	0.036	0.040	0.045	0.053	0.039	0.050	0.081	0.104	0.050	0.096	0.152	0.226
	DICT 3-Item	0.326	0.338	0.368	0.397	0.341	0.416	0.511	0.569	0.409	0.615	0.665	0.809
	DICT 4-Item	0.274	0.268	0.314	0.322	0.269	0.363	0.440	0.532	0.352	0.499	0.611	0.786
	DICT 5-Item	0.234	0.248	0.249	0.310	0.256	0.300	0.419	0.480	0.320	0.445	0.608	0.678
	SSC 3-Item	0.150	0.157	0.189	0.223	0.171	0.207	0.281	0.371	0.197	0.334	0.495	0.576
	SSC 4-Item	0.163	0.158	0.176	0.196	0.170	0.213	0.269	0.372	0.188	0.312	0.473	0.604
	SSC 5-Item	0.101	0.106	0.124	0.147	0.107	0.159	0.207	0.269	0.150	0.240	0.339	0.462
0.10	ICT 3-Item	0.078	0.098	0.125	0.166	0.098	0.156	0.243	0.348	0.138	0.323	0.450	0.613
	ICT 4-Item	0.076	0.091	0.119	0.153	0.093	0.154	0.240	0.313	0.137	0.289	0.486	0.588
	ICT 5-Item	0.064	0.075	0.091	0.130	0.069	0.121	0.210	0.291	0.108	0.240	0.384	0.511
	DICT 3-Item	0.456	0.494	0.565	0.696	0.573	0.596	0.774	0.870	0.605	0.838	0.917	0.947
	DICT 4-Item	0.404	0.443	0.542	0.611	0.446	0.609	0.736	0.836	0.570	0.738	0.883	0.902
	DICT 5-Item	0.346	0.398	0.449	0.559	0.407	0.549	0.706	0.788	0.525	0.761	0.857	0.908
	SSC 3-Item	0.228	0.299	0.351	0.437	0.289	0.414	0.550	0.651	0.392	0.598	0.786	0.869
	SSC 4-Item	0.244	0.272	0.322	0.431	0.292	0.399	0.539	0.648	0.410	0.599	0.775	0.816
	SSC 5-Item	0.171	0.204	0.275	0.298	0.202	0.307	0.434	0.589	0.284	0.492	0.649	0.795

Table 10 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.15	ICT 3-Item	0.098	0.141	0.211	0.327	0.141	0.276	0.386	0.524	0.225	0.488	0.667	0.809
	ICT 4-Item	0.103	0.142	0.208	0.282	0.138	0.254	0.406	0.540	0.207	0.438	0.630	0.748
	ICT 5-Item	0.084	0.109	0.163	0.224	0.116	0.222	0.351	0.434	0.172	0.411	0.589	0.723
	DICT 3-Item	0.553	0.650	0.761	0.832	0.617	0.783	0.879	0.947	0.797	0.898	0.930	0.987
	DICT 4-Item	0.486	0.594	0.666	0.750	0.604	0.736	0.834	0.923	0.691	0.872	0.951	0.976
	DICT 5-Item	0.430	0.529	0.628	0.684	0.514	0.655	0.819	0.862	0.672	0.846	0.907	0.976
	SSC 3-Item	0.323	0.406	0.471	0.609	0.391	0.578	0.715	0.808	0.547	0.753	0.868	0.913
	SSC 4-Item	0.316	0.373	0.563	0.629	0.380	0.574	0.676	0.806	0.584	0.765	0.849	0.928
	SSC 5-Item	0.247	0.294	0.390	0.498	0.296	0.455	0.606	0.768	0.444	0.645	0.827	0.875
0.20	ICT 3-Item	0.126	0.212	0.297	0.394	0.173	0.352	0.525	0.651	0.331	0.595	0.721	0.889
	ICT 4-Item	0.129	0.181	0.301	0.354	0.198	0.342	0.485	0.634	0.297	0.620	0.776	0.881
	ICT 5-Item	0.114	0.167	0.237	0.324	0.154	0.293	0.457	0.586	0.264	0.500	0.685	0.817
	DICT 3-Item	0.647	0.732	0.794	0.879	0.745	0.846	0.942	0.957	0.843	0.892	0.982	0.995
	DICT 4-Item	0.548	0.662	0.733	0.800	0.664	0.815	0.920	0.948	0.795	0.884	0.972	0.996
	DICT 5-Item	0.527	0.641	0.705	0.798	0.661	0.818	0.900	0.908	0.735	0.877	0.958	0.979
	SSC 3-Item	0.438	0.516	0.620	0.714	0.467	0.739	0.828	0.887	0.707	0.827	0.926	0.956
	SSC 4-Item	0.378	0.507	0.605	0.718	0.507	0.737	0.834	0.872	0.669	0.864	0.939	0.970
	SSC 5-Item	0.310	0.372	0.498	0.602	0.365	0.545	0.724	0.849	0.585	0.811	0.885	0.926

Table 10 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.25	ICT 3-Item	0.161	0.239	0.382	0.506	0.245	0.476	0.617	0.752	0.417	0.652	0.871	0.868
	ICT 4-Item	0.154	0.250	0.350	0.498	0.236	0.443	0.648	0.786	0.414	0.681	0.823	0.903
	ICT 5-Item	0.123	0.220	0.324	0.445	0.206	0.424	0.582	0.652	0.372	0.639	0.815	0.882
	DICT 3-Item	0.729	0.831	0.893	0.908	0.753	0.910	0.907	0.982	0.868	0.968	0.981	0.991
	DICT 4-Item	0.652	0.730	0.820	0.869	0.695	0.884	0.912	0.975	0.836	0.956	0.971	0.977
	DICT 5-Item	0.590	0.690	0.835	0.906	0.703	0.867	0.937	0.967	0.875	0.936	0.965	0.983
	SSC 3-Item	0.450	0.518	0.718	0.833	0.596	0.734	0.868	0.913	0.724	0.909	0.936	0.967
	SSC 4-Item	0.426	0.615	0.688	0.781	0.532	0.796	0.869	0.886	0.736	0.876	0.954	0.979
	SSC 5-Item	0.332	0.469	0.645	0.680	0.479	0.682	0.840	0.863	0.638	0.899	0.940	0.958
0.35	ICT 3-Item	0.223	0.356	0.546	0.599	0.358	0.677	0.780	0.854	0.558	0.832	0.907	0.950
	ICT 4-Item	0.191	0.352	0.500	0.602	0.318	0.591	0.730	0.828	0.551	0.776	0.931	0.925
	ICT 5-Item	0.175	0.320	0.465	0.608	0.264	0.530	0.718	0.795	0.475	0.791	0.863	0.919
	DICT 3-Item	0.811	0.919	0.917	0.970	0.926	0.941	0.980	0.989	0.963	1.008	0.994	0.993
	DICT 4-Item	0.731	0.813	0.931	0.911	0.820	0.926	0.944	0.972	0.941	0.965	0.986	1.001
	DICT 5-Item	0.652	0.841	0.896	0.946	0.801	0.904	0.949	0.986	0.902	0.983	0.996	0.989
	SSC 3-Item	0.562	0.690	0.813	0.871	0.652	0.822	0.897	0.967	0.839	0.964	0.971	0.975
	SSC 4-Item	0.558	0.695	0.844	0.917	0.714	0.857	0.959	0.976	0.863	0.926	0.983	0.977
	SSC 5-Item	0.413	0.619	0.753	0.817	0.553	0.785	0.873	0.942	0.773	0.931	0.966	0.981

Table 10 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.45	ICT 3-Item	0.270	0.490	0.658	0.722	0.431	0.689	0.842	0.909	0.664	0.900	0.933	0.963
	ICT 4-Item	0.233	0.445	0.633	0.755	0.442	0.720	0.825	0.907	0.656	0.862	0.939	0.980
	ICT 5-Item	0.231	0.398	0.586	0.659	0.353	0.627	0.809	0.855	0.570	0.841	0.937	0.975
	DICT 3-Item	0.843	0.918	0.946	0.985	0.917	0.964	0.957	0.978	0.961	0.972	0.994	1.007
	DICT 4-Item	0.812	0.928	0.923	0.945	0.856	0.950	0.971	0.994	0.923	0.989	0.992	1.002
	DICT 5-Item	0.804	0.879	0.919	0.939	0.880	0.981	0.960	0.989	0.953	0.959	0.987	1.004
	SSC 3-Item	0.666	0.783	0.889	0.951	0.785	0.909	0.958	0.975	0.895	0.966	0.979	0.982
	SSC 4-Item	0.596	0.789	0.881	0.928	0.805	0.922	0.970	0.986	0.853	0.965	1.000	0.988
	SSC 5-Item	0.519	0.695	0.898	0.889	0.692	0.840	0.907	0.945	0.799	0.948	0.975	0.978

*The SSC assumption of using the normal approximation of a binomial for $\boldsymbol{\pi_s}$ =.01 was not met.

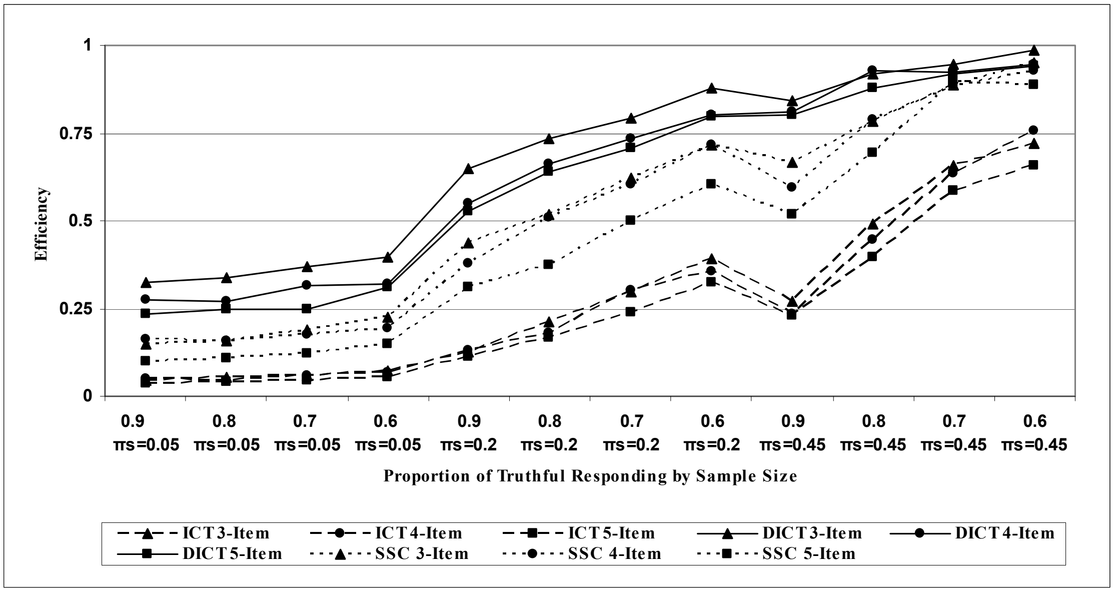


Figure 13. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response models to the Unrelated Question Technique (optimal), *n* = 150

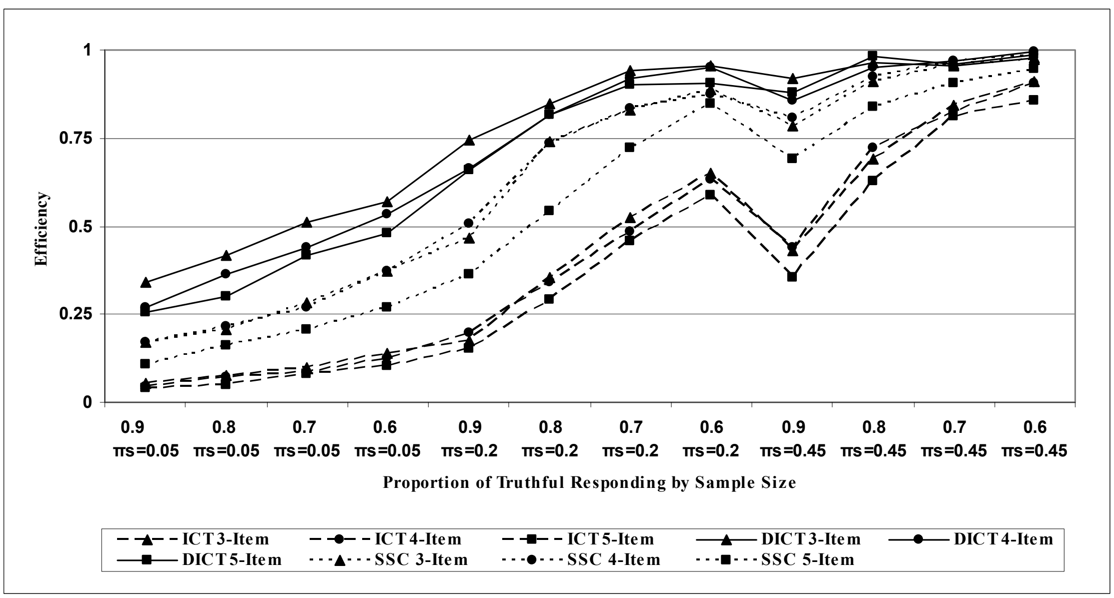


Figure 14. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response models to the Unrelated Question Technique (optimal), *n* = 500

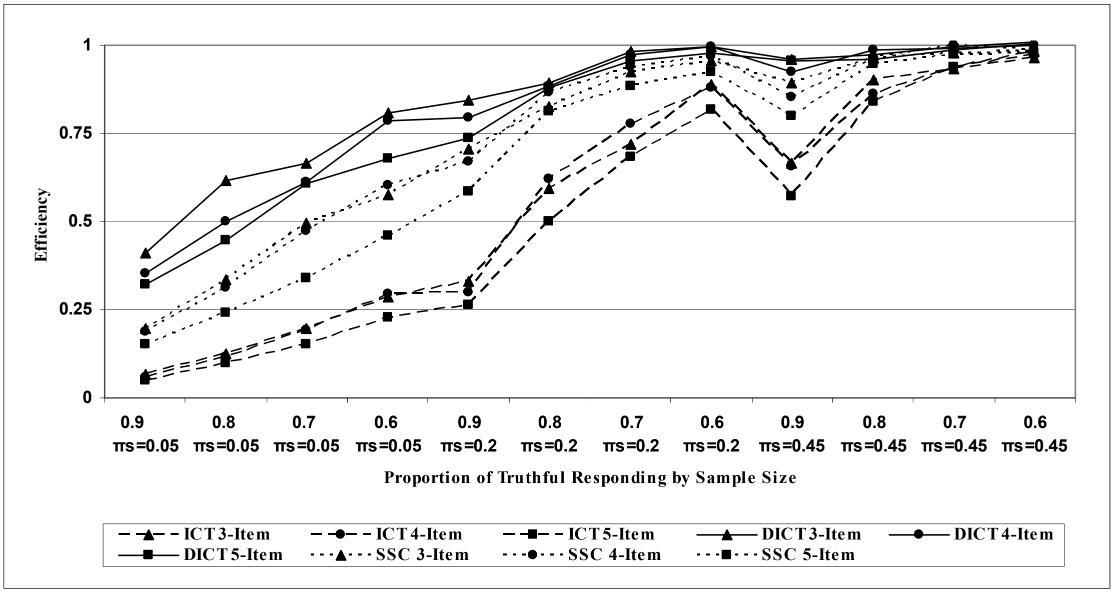


Figure 15. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response models to the Unrelated Question Technique (optimal), *n* = 1,500

When π_s was larger (i.e., 0.25, 0.35, and 0.45), as π_s increased the double item count technique (DICT) optimal models were nearly as efficient as the unrelated question technique (UQT) optimal for small sample sizes (*n* = 150) in the presence of higher non-compliance rates (i.e., percent of truthful reporting $\leq$ 0.70). For mid level sample sizes (*n* = 500), as π_s increased the DICT optimal models were nearly as efficient as the UQT optimal in the presence of moderate non-compliance rates (i.e., percent of truthful reporting $\leq$ 0.80). When sample sizes increased to 1,500, all DICT optimal models approached the efficiency of the UQT across all non-compliant rates (i.e., percent of truthful reporting $\leq$ 0.90). The optimal SSC 3-item model proved to be nearly as efficient as the UQT optimal for small sample sizes (i.e., *n* = 150) in the presence of high non-compliance (i.e., percent of truthful reporting = 0.60) as π_s increased. For moderate samples sizes (*n* = 500) the SSC 3-item and 4-item optimal models were nearly as efficient as the UQT in the presence of moderate non-compliance rates (i.e., percent of truthful reporting $\leq$ 0.80). This trend continued for all SSC optimal models when sample sizes were large (*n* = 1,500). For the optimal ICT, as π_s increased, the model was nearly as efficient as the UQT optimal for mid-sample sizes (*n* = 500) when non-compliance was high (i.e., percent of truthful reporting = 0.60) and for large sample sizes (n=1,500) when non-compliance was also high (i.e., percent of truthful reporting $\leq$ 0.70).

As is evident, in the case of comparing the UQT optimal to the non-random response (NRR) optimal models, a pattern was apparent. Across all sample sizes, sensitive prevalent rates and equivalent non-compliance rates, the DICT optimal models were the most efficient of the NRR optimal models since the ratio of MSE as compared with the UQT optimal was greater than any of the corresponding ratio of MSE of the single sample count technique (SSC) and item count technique (ICT) optimal models. As is seen in each figure, the double item count technique (DICT) optimal 3-item list was most efficient followed by the DICT optimal 4-item and 5-Item list models. Only in the presence of high non-compliance (i.e., percent of truthful reporting $\geq$ 0.70), larger sample sizes (*n* = 1,500) and larger sensitive prevalent rates (π_s = 0.45) did the NRR optimal models and UQT optimal approach unity.

Unrelated Question Technique (Practical) vs. Non-random Response Techniques (Optimal)

Table 11 displays the results of the non-compliance study, comparing the practical UQT technique to each of the NRR techniques in the presence of non-compliance. Figures 16, 17, and 18 display the ratio of MSE by proportion of truthful reporting (0.90, 0.80, 0.70, and 0.60) and sample size (150, 500, and 1,500) for sensitive prevalent rates

π_s = 0.05 (small), π_s = 0.20 (medium), and π_s = 0.45 (large). In each figure, the plot of the MSE by NRR technique is coded by line and plot type as was described in the previous section. For these analyses, the study continued to assume that a ratio of MSE close to unity implied that the comparing models were similar in their efficiency.

As is evident in the table and figures, when π_s is small (i.e., 0.01, 0.03, and 0.05) across all sample sizes and non-compliance rates, the UQT practical model was *less* efficient than all NRR models. When π_s was moderate (i.e., 0.10, 0.15, and 0.20), this trend continued where the UQT practical model was generally *less* efficient compared to all NRR optimal models. For larger π_s (i.e., 0.25, 0.35, and 0.45), the UQT practical was generally *less* efficient compared with all NRR models except in the cases where sample sizes were large (n = 1,500) and non-compliance was high (i.e., percent of truthful reporting = 0.60). In these cases, the UQT practical was nearly as efficient as all NRR optimal models.

Table 11

Study of Non-compliance: Unrelated Question Technique Practical vs. Non-random Response Optimal Models

π_s	Non-random Response Model	*n* = 150				*n* = 500				*n* = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.01	ICT 3-Item	1.980	2.074	2.041	1.826	1.916	1.893	1.900	1.810	1.846	1.828	2.049	1.974
	ICT 4-Item	1.868	1.789	1.778	1.659	1.682	1.716	1.708	1.756	1.768	1.746	1.841	1.758
	ICT 5-Item	1.299	1.421	1.320	1.404	1.479	1.430	1.421	1.315	1.334	1.365	1.438	1.388
	DICT 3-Item	13.488	14.008	13.628	13.952	12.991	13.109	13.100	13.470	13.837	12.957	13.020	11.425
	DICT 4-Item	11.341	11.052	10.908	11.963	10.687	11.096	10.956	11.137	11.486	11.080	10.174	10.021
	DICT 5-Item	8.917	9.586	9.202	9.426	9.059	9.737	10.453	8.958	9.181	9.595	9.535	10.275
	SSC 3-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 4-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 5-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.03	ICT 3-Item	1.933	1.923	1.808	1.789	1.936	1.990	1.951	1.924	2.012	1.922	1.767	1.787
	ICT 4-Item	1.745	1.729	1.758	1.829	1.832	1.771	1.690	1.698	1.762	1.707	1.809	1.610
	ICT 5-Item	1.434	1.340	1.359	1.470	1.476	1.435	1.486	1.377	1.386	1.339	1.437	1.434
	DICT 3-Item	13.547	13.197	12.693	11.863	13.013	13.049	12.447	11.584	12.843	10.539	9.448	8.442
	DICT 4-Item	10.320	10.608	10.989	10.619	10.257	10.482	9.671	8.593	10.900	9.960	7.901	6.350
	DICT 5-Item	9.884	9.252	9.602	9.487	9.865	9.121	9.840	8.117	9.366	8.683	6.854	6.044
	SSC 3-Item	6.436	6.675	6.157	6.177	5.389	6.583	6.122	5.474	6.380	5.972	5.484	4.609
	SSC 4-Item	6.484	6.372	5.846	6.307	6.538	6.297	6.012	5.698	6.102	5.745	5.504	4.724
	SSC 5-Item	4.079	4.069	4.206	3.863	4.278	4.229	3.780	3.802	4.119	4.090	3.931	3.578

Table 11 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.05	ICT 3-Item	1.879	2.058	1.934	1.990	1.962	2.094	1.878	1.868	2.032	2.014	1.807	1.685
	ICT 4-Item	1.970	1.815	1.986	1.928	1.715	1.999	1.672	1.686	1.784	1.849	1.750	1.747
	ICT 5-Item	1.505	1.552	1.491	1.465	1.460	1.405	1.554	1.395	1.490	1.534	1.390	1.337
	DICT 3-Item	13.455	13.025	12.296	11.017	12.795	11.605	9.843	7.658	12.153	9.779	6.073	4.794
	DICT 4-Item	11.316	10.319	10.492	8.925	10.075	10.116	8.486	7.158	10.480	7.938	5.585	4.658
	DICT 5-Item	9.678	9.556	8.329	8.594	9.580	8.359	8.074	6.453	9.524	7.071	5.557	4.017
	SSC 3-Item	6.201	6.035	6.314	6.191	6.405	5.781	5.425	4.990	5.869	5.305	4.522	3.410
	SSC 4-Item	6.744	6.074	5.881	5.430	6.385	5.950	5.178	5.006	5.591	4.967	4.325	3.576
	SSC 5-Item	4.163	4.092	4.129	4.091	4.023	4.446	3.986	3.613	4.469	3.817	3.099	2.736
0.10	ICT 3-Item	2.063	2.087	1.951	1.884	2.108	1.938	1.794	1.702	1.969	1.960	1.550	1.471
	ICT 4-Item	2.006	1.939	1.858	1.741	2.001	1.907	1.772	1.533	1.959	1.753	1.674	1.409
	ICT 5-Item	1.709	1.585	1.420	1.478	1.483	1.499	1.550	1.424	1.542	1.452	1.322	1.226
	DICT 3-Item	12.113	10.511	8.835	7.893	12.372	7.385	5.717	4.255	8.615	5.077	3.158	2.271
	DICT 4-Item	10.717	9.427	8.473	6.938	9.640	7.550	5.433	4.087	8.122	4.470	3.043	2.163
	DICT 5-Item	9.181	8.475	7.021	6.339	8.791	6.803	5.212	3.855	7.485	4.610	2.952	2.178
	SSC 3-Item	6.055	6.365	5.482	4.954	6.238	5.127	4.064	3.184	5.589	3.626	2.709	2.084
	SSC 4-Item	6.469	5.782	5.029	4.896	6.306	4.950	3.979	3.171	5.837	3.632	2.669	1.958
	SSC 5-Item	4.547	4.348	4.297	3.380	4.363	3.809	3.204	2.882	4.053	2.979	2.234	1.907

Table 11 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.15	ICT 3-Item	1.981	2.018	2.000	2.136	2.118	2.049	1.658	1.532	1.992	1.755	1.467	1.352
	ICT 4-Item	2.095	2.022	1.973	1.846	2.063	1.888	1.743	1.580	1.829	1.576	1.386	1.250
	ICT 5-Item	1.705	1.561	1.545	1.462	1.743	1.651	1.506	1.269	1.526	1.480	1.297	1.208
	DICT 3-Item	11.214	9.282	7.227	5.439	9.256	5.811	3.780	2.768	7.051	3.229	2.046	1.649
	DICT 4-Item	9.855	8.481	6.322	4.905	9.054	5.463	3.582	2.700	6.113	3.136	2.092	1.631
	DICT 5-Item	8.709	7.544	5.966	4.474	7.713	4.865	3.521	2.519	5.944	3.042	1.996	1.631
	SSC 3-Item	6.545	5.799	4.472	3.983	5.863	4.292	3.075	2.363	4.845	2.708	1.910	1.526
	SSC 4-Item	6.411	5.329	5.351	4.112	5.698	4.259	2.907	2.358	5.168	2.753	1.867	1.551
	SSC 5-Item	5.008	4.195	3.703	3.256	4.445	3.374	2.606	2.245	3.933	2.321	1.820	1.462
0.20	ICT 3-Item	2.102	2.236	1.973	1.771	1.967	1.831	1.600	1.411	2.080	1.553	1.244	1.246
	ICT 4-Item	2.146	1.907	1.998	1.593	2.258	1.779	1.479	1.373	1.870	1.620	1.340	1.235
	ICT 5-Item	1.903	1.765	1.570	1.456	1.751	1.521	1.394	1.269	1.658	1.305	1.182	1.146
	DICT 3-Item	10.796	7.734	5.270	3.954	8.489	4.397	2.872	2.074	5.305	2.329	1.696	1.395
	DICT 4-Item	9.147	6.996	4.863	3.600	7.567	4.233	2.806	2.056	5.001	2.308	1.678	1.396
	DICT 5-Item	8.798	6.775	4.681	3.591	7.529	4.249	2.745	1.968	4.622	2.291	1.654	1.373
	SSC 3-Item	7.315	5.450	4.117	3.214	5.324	3.838	2.525	1.923	4.443	2.159	1.599	1.341
	SSC 4-Item	6.310	5.354	4.012	3.231	5.774	3.826	2.544	1.890	4.206	2.257	1.621	1.359
	SSC 5-Item	5.170	3.934	3.308	2.710	4.160	2.829	2.206	1.841	3.679	2.118	1.528	1.298

Table 11 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.25	ICT 3-Item	2.303	1.984	1.932	1.742	2.236	1.899	1.489	1.349	2.024	1.378	1.302	1.103
	ICT 4-Item	2.202	2.078	1.770	1.715	2.156	1.768	1.566	1.410	2.010	1.439	1.230	1.148
	ICT 5-Item	1.765	1.832	1.641	1.532	1.884	1.693	1.406	1.170	1.804	1.350	1.218	1.121
	DICT 3-Item	10.437	6.904	4.515	3.124	6.877	3.633	2.190	1.762	4.210	2.045	1.466	1.260
	DICT 4-Item	9.329	6.069	4.150	2.989	6.341	3.526	2.203	1.750	4.052	2.020	1.451	1.242
	DICT 5-Item	8.436	5.737	4.222	3.118	6.420	3.458	2.264	1.735	4.243	1.977	1.443	1.250
	SSC 3-Item	6.444	4.305	3.633	2.867	5.444	2.928	2.096	1.637	3.509	1.921	1.399	1.230
	SSC 4-Item	6.090	5.109	3.478	2.686	4.857	3.177	2.100	1.590	3.572	1.852	1.426	1.245
	SSC 5-Item	4.743	3.899	3.263	2.338	4.371	2.723	2.028	1.548	3.094	1.899	1.405	1.218
0.35	ICT 3-Item	2.511	2.046	1.878	1.445	2.317	1.878	1.412	1.235	1.867	1.362	1.160	1.094
	ICT 4-Item	2.146	2.029	1.722	1.453	2.053	1.639	1.321	1.198	1.842	1.269	1.190	1.065
	ICT 5-Item	1.965	1.842	1.599	1.465	1.708	1.472	1.299	1.150	1.588	1.295	1.104	1.058
	DICT 3-Item	9.125	5.288	3.154	2.338	5.991	2.611	1.774	1.431	3.220	1.648	1.271	1.143
	DICT 4-Item	8.220	4.677	3.203	2.196	5.299	2.569	1.709	1.407	3.144	1.579	1.261	1.153
	DICT 5-Item	7.340	4.840	3.083	2.281	5.178	2.509	1.719	1.427	3.015	1.607	1.274	1.139
	SSC 3-Item	6.326	3.972	2.797	2.101	4.217	2.280	1.623	1.400	2.805	1.578	1.241	1.123
	SSC 4-Item	6.283	4.001	2.902	2.212	4.614	2.379	1.736	1.411	2.884	1.515	1.257	1.125
	SSC 5-Item	4.652	3.565	2.590	1.970	3.573	2.177	1.579	1.363	2.585	1.523	1.235	1.129

Table 11 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.45	ICT 3-Item	2.513	2.146	1.746	1.397	2.141	1.509	1.292	1.174	1.718	1.276	1.103	1.058
	ICT 4-Item	2.164	1.947	1.679	1.462	2.191	1.577	1.264	1.171	1.699	1.222	1.109	1.077
	ICT 5-Item	2.146	1.741	1.554	1.276	1.752	1.374	1.241	1.105	1.477	1.192	1.108	1.071
	DICT 3-Item	7.839	4.017	2.509	1.907	4.552	2.110	1.466	1.264	2.488	1.379	1.175	1.106
	DICT 4-Item	7.553	4.064	2.448	1.829	4.247	2.080	1.488	1.284	2.390	1.402	1.172	1.101
	DICT 5-Item	7.470	3.849	2.437	1.818	4.365	2.149	1.471	1.277	2.468	1.360	1.166	1.103
	SSC 3-Item	6.192	3.429	2.357	1.841	3.894	1.992	1.469	1.260	2.317	1.369	1.156	1.078
	SSC 4-Item	5.541	3.453	2.337	1.797	3.994	2.020	1.486	1.274	2.208	1.367	1.182	1.085
	SSC 5-Item	4.829	3.042	2.381	1.720	3.431	1.839	1.391	1.221	2.069	1.344	1.152	1.074

*The SSC assumption of using the normal approximation of a binomial for $\boldsymbol{\pi_s}$ =.01 was not met.

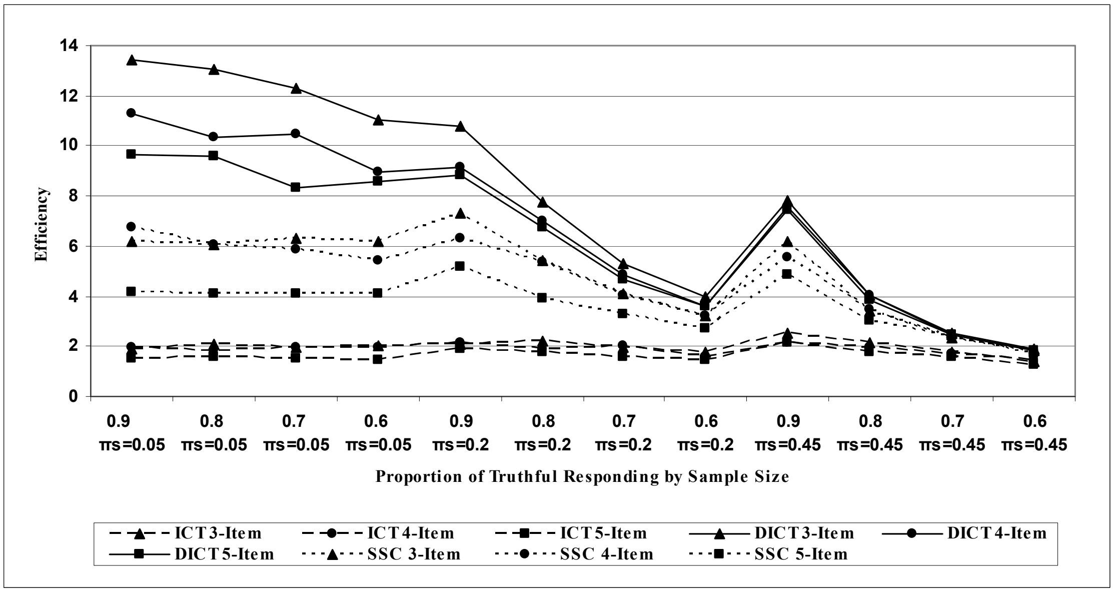


Figure 16. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response Models to the Unrelated Question Technique (Practical), $n = 150$

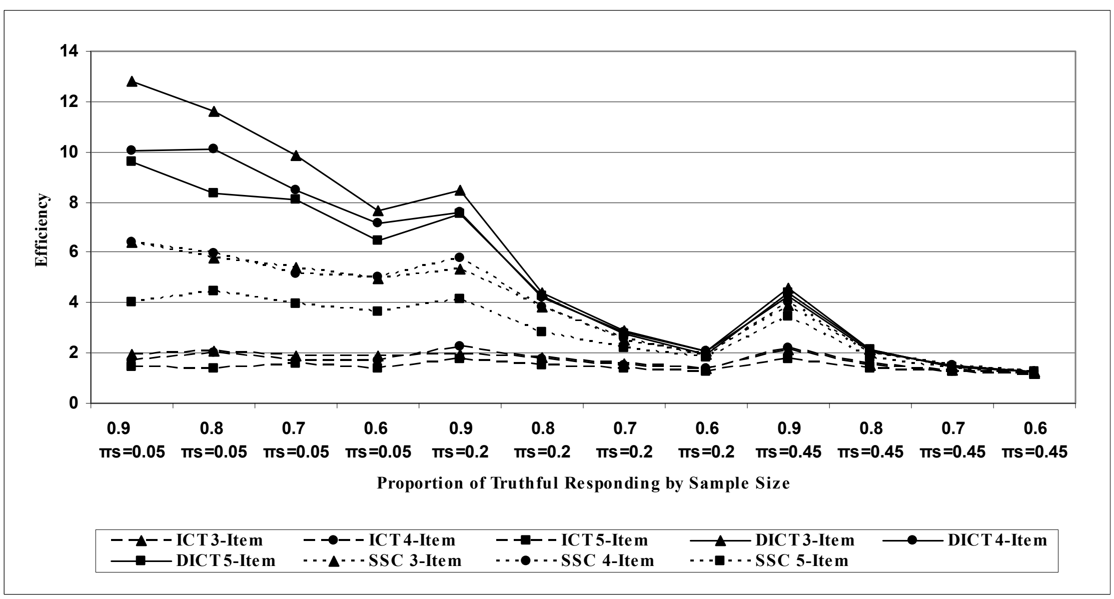


Figure 17. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response Models to the Unrelated Question Technique (Practical), $n = 500$

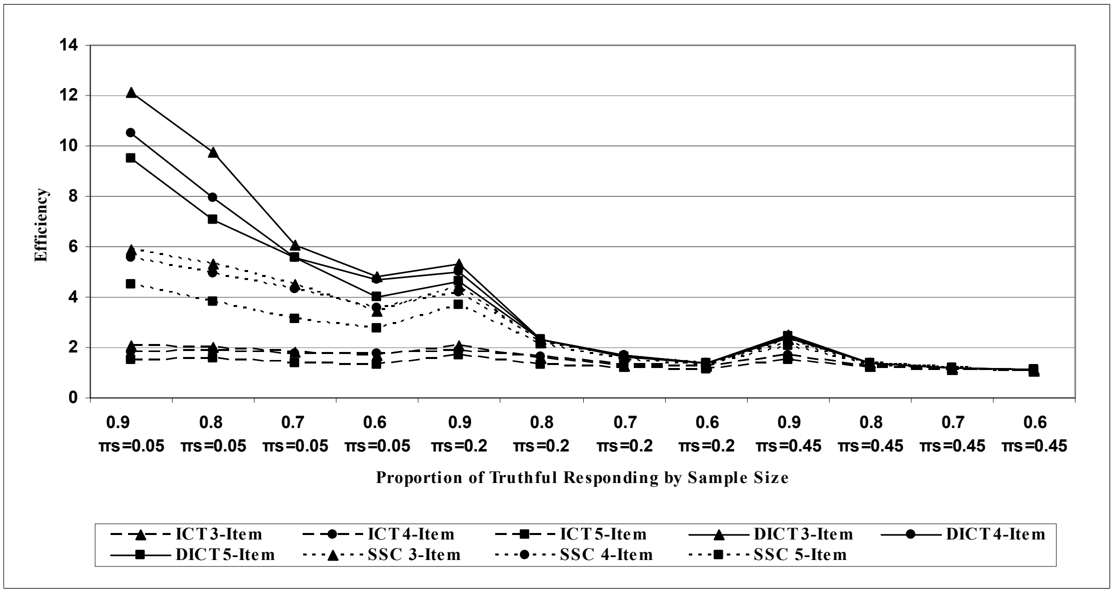


Figure 18. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response Models to the Unrelated Question Technique (Practical), *n* = 1,500

As is evident, in the case of comparing the UQT practical to the non-random response (NRR) optimal models, a similar pattern to the previous analysis was also apparent. Across all sample sizes, sensitive prevalent rates and equivalent non-compliance rates, the double item count technique (DICT) optimal models were the most efficient of the NRR optimal models since the ratio of MSE as compared with the UQT practical was greater than any of the corresponding ratio of MSE of the single sample count technique (SSC) and item count technique (ICT) optimal models. As is seen in each figure, the DICT optimal 3-item list was most efficient followed by the DICT optimal 4-item and 5-Item list models. Only in the presence of high non-compliance (i.e., percent of truthful reporting $\leq$ 0.70), larger sample sizes (*n* = 1,500) and larger sensitive prevalent rates (π_S =0.45) did the NRR optimal models and UQT practical approach unity.

**Forced Choice Technique vs.
Non-random Response
(Optimal)**

Table 12 displays the results of the non-compliance study, comparing the optimal UQT technique to each of the NRR optimal techniques. Figures 19, 20, and 21 display the ratio of MSE by the proportion of truthful reporting (0.90, 0.80, 0.70, and 0.60) and sample size for sensitive prevalent rates $\pi_s = 0.05$ (small), $\pi_s = 0.20$ (medium), and $\pi_s = 0.45$ (large). In each figure, the plot of the MSE by NRR technique is coded by line and plot type as was described in previous sections. For these analyses, the study continued to assume that a ratio of MSE close to unity implied that the models were similar in their efficiency.

As is evident in the table and figures, when π_s was small (i.e., 0.01, 0.03, and 0.05) across all sample sizes and non-compliance rates, the double item count technique (DICT) 3-Item optimal model was generally *as* or more efficient as the FCT model; whereas DICT 4-Item, 5-Item and optimal SSC and ICT models were *less* efficient then the FCT.

When π_s was moderate (i.e., 0.10, 0.15, and 0.20), as π_s and non-compliance rates increased, the FCT approached the efficiency of the DICT 3-item optimal model across all sample sizes and non-compliance rates. For the optimal SSC models, when sample sizes were large (n = 1,500) and non-compliance was moderate to high (i.e., percent of truthful reporting ≤ 0.70), the 3-item and 4-item optimal models were nearly as efficient as the FCT for increasing π_s. The optimal ICT models generally were *less* efficient compared to the FCT.

Table 12

Study of Non-compliance: Forced Choice Technique vs. Non-random Response Optimal Models

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.01	ICT 3-Item	0.162	0.169	0.167	0.150	0.157	0.155	0.157	0.152	0.151	0.152	0.175	0.175
	ICT 4-Item	0.152	0.146	0.146	0.136	0.137	0.141	0.141	0.147	0.145	0.145	0.157	0.156
	ICT 5-Item	0.106	0.116	0.108	0.115	0.121	0.117	0.118	0.110	0.109	0.114	0.123	0.123
	DICT 3-Item	1.100	1.144	1.116	1.146	1.061	1.076	1.085	1.129	1.134	1.079	1.113	1.011
	DICT 4-Item	0.925	0.903	0.893	0.983	0.873	0.911	0.907	0.933	0.942	0.923	0.869	0.887
	DICT 5-Item	0.727	0.783	0.753	0.774	0.740	0.799	0.866	0.751	0.753	0.799	0.815	0.909
	SSC 3-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 4-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SSC 5-Item*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.03	ICT 3-Item	0.159	0.161	0.154	0.157	0.161	0.173	0.182	0.196	0.173	0.186	0.203	0.250
	ICT 4-Item	0.144	0.144	0.150	0.161	0.153	0.154	0.158	0.173	0.151	0.165	0.208	0.225
	ICT 5-Item	0.118	0.112	0.116	0.129	0.123	0.125	0.139	0.140	0.119	0.130	0.165	0.200
	DICT 3-Item	1.116	1.102	1.083	1.043	1.084	1.135	1.160	1.181	1.102	1.021	1.087	1.179
	DICT 4-Item	0.850	0.885	0.937	0.934	0.854	0.912	0.902	0.876	0.935	0.965	0.909	0.887
	DICT 5-Item	0.814	0.772	0.819	0.834	0.822	0.793	0.917	0.827	0.803	0.841	0.789	0.844
	SSC 3-Item	0.530	0.557	0.525	0.543	0.449	0.573	0.571	0.558	0.547	0.579	0.631	0.644
	SSC 4-Item	0.534	0.532	0.499	0.555	0.544	0.548	0.560	0.581	0.523	0.557	0.633	0.660
	SSC 5-Item	0.336	0.340	0.359	0.340	0.356	0.368	0.352	0.388	0.353	0.396	0.452	0.500

Table 12 (continued)

π_s	Non-random Response Model	*n* = 150				*n* = 500				*n* = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.05	ICT 3-Item	0.157	0.178	0.177	0.196	0.168	0.200	0.210	0.251	0.187	0.244	0.300	0.375
	ICT 4-Item	0.164	0.157	0.181	0.190	0.147	0.191	0.187	0.226	0.165	0.224	0.290	0.389
	ICT 5-Item	0.126	0.134	0.136	0.144	0.125	0.134	0.174	0.187	0.137	0.186	0.230	0.297
	DICT 3-Item	1.122	1.124	1.123	1.084	1.097	1.109	1.101	1.028	1.121	1.183	1.007	1.067
	DICT 4-Item	0.944	0.891	0.958	0.878	0.864	0.967	0.949	0.961	0.967	0.960	0.926	1.037
	DICT 5-Item	0.807	0.825	0.761	0.846	0.821	0.799	0.903	0.866	0.879	0.855	0.922	0.894
	SSC 3-Item	0.517	0.521	0.577	0.609	0.549	0.552	0.607	0.670	0.541	0.642	0.750	0.759
	SSC 4-Item	0.562	0.524	0.537	0.534	0.547	0.569	0.579	0.672	0.516	0.601	0.717	0.796
	SSC 5-Item	0.347	0.353	0.377	0.402	0.345	0.425	0.446	0.485	0.412	0.462	0.514	0.609
0.10	ICT 3-Item	0.178	0.203	0.225	0.264	0.200	0.250	0.328	0.425	0.231	0.410	0.510	0.662
	ICT 4-Item	0.174	0.189	0.214	0.244	0.190	0.246	0.324	0.383	0.230	0.367	0.551	0.634
	ICT 5-Item	0.148	0.154	0.164	0.207	0.141	0.194	0.283	0.356	0.181	0.304	0.435	0.552
	DICT 3-Item	1.048	1.023	1.018	1.106	1.172	0.954	1.045	1.063	1.012	1.061	1.040	1.023
	DICT 4-Item	0.927	0.917	0.977	0.972	0.913	0.975	0.993	1.021	0.954	0.935	1.002	0.974
	DICT 5-Item	0.794	0.825	0.809	0.888	0.833	0.879	0.953	0.963	0.879	0.964	0.972	0.981
	SSC 3-Item	0.524	0.619	0.632	0.694	0.591	0.662	0.743	0.796	0.656	0.758	0.892	0.938
	SSC 4-Item	0.560	0.562	0.580	0.686	0.597	0.640	0.727	0.792	0.685	0.759	0.879	0.882
	SSC 5-Item	0.393	0.423	0.495	0.474	0.413	0.492	0.586	0.720	0.476	0.623	0.735	0.859

Table 12 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.15	ICT 3-Item	0.179	0.227	0.297	0.418	0.227	0.357	0.447	0.575	0.301	0.546	0.705	0.836
	ICT 4-Item	0.189	0.227	0.293	0.362	0.221	0.329	0.470	0.593	0.277	0.490	0.666	0.773
	ICT 5-Item	0.154	0.175	0.229	0.287	0.186	0.287	0.406	0.476	0.231	0.460	0.623	0.748
	DICT 3-Item	1.013	1.043	1.072	1.066	0.990	1.012	1.019	1.039	1.067	1.004	0.984	1.020
	DICT 4-Item	0.891	0.953	0.937	0.961	0.969	0.951	0.966	1.014	0.925	0.975	1.006	1.009
	DICT 5-Item	0.787	0.848	0.885	0.877	0.825	0.847	0.949	0.946	0.899	0.946	0.960	1.009
	SSC 3-Item	0.591	0.652	0.663	0.780	0.627	0.747	0.829	0.887	0.733	0.842	0.919	0.944
	SSC 4-Item	0.579	0.599	0.793	0.806	0.610	0.741	0.784	0.885	0.782	0.856	0.898	0.959
	SSC 5-Item	0.453	0.472	0.549	0.638	0.475	0.587	0.703	0.843	0.595	0.722	0.875	0.905
0.20	ICT 3-Item	0.199	0.292	0.368	0.456	0.239	0.411	0.570	0.686	0.396	0.633	0.743	0.905
	ICT 4-Item	0.203	0.249	0.373	0.410	0.275	0.399	0.527	0.667	0.356	0.660	0.800	0.897
	ICT 5-Item	0.180	0.230	0.293	0.375	0.213	0.342	0.497	0.617	0.315	0.532	0.706	0.832
	DICT 3-Item	1.023	1.010	0.984	1.018	1.032	0.987	1.024	1.008	1.009	0.949	1.013	1.014
	DICT 4-Item	0.867	0.913	0.908	0.927	0.920	0.950	1.000	0.999	0.951	0.940	1.002	1.014
	DICT 5-Item	0.834	0.884	0.874	0.925	0.916	0.954	0.979	0.956	0.879	0.933	0.987	0.997
	SSC 3-Item	0.693	0.712	0.769	0.828	0.647	0.862	0.900	0.934	0.845	0.880	0.954	0.974
	SSC 4-Item	0.598	0.699	0.749	0.832	0.702	0.859	0.907	0.919	0.800	0.920	0.968	0.987
	SSC 5-Item	0.490	0.514	0.618	0.698	0.506	0.635	0.786	0.895	0.700	0.863	0.912	0.943

Table 12 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.25	ICT 3-Item	0.230	0.299	0.440	0.557	0.309	0.525	0.649	0.776	0.469	0.677	0.887	0.877
	ICT 4-Item	0.220	0.314	0.404	0.548	0.298	0.489	0.683	0.811	0.466	0.708	0.839	0.913
	ICT 5-Item	0.176	0.276	0.374	0.490	0.260	0.468	0.613	0.673	0.418	0.664	0.830	0.891
	DICT 3-Item	1.041	1.042	1.029	0.999	0.950	1.005	0.955	1.014	0.975	1.005	0.999	1.002
	DICT 4-Item	0.930	0.916	0.946	0.956	0.876	0.976	0.961	1.007	0.939	0.993	0.989	0.988
	DICT 5-Item	0.841	0.866	0.962	0.997	0.887	0.957	0.987	0.998	0.983	0.972	0.983	0.994
	SSC 3-Item	0.643	0.650	0.828	0.917	0.752	0.810	0.914	0.942	0.813	0.945	0.953	0.978
	SSC 4-Item	0.607	0.771	0.793	0.859	0.671	0.879	0.916	0.915	0.827	0.910	0.972	0.990
	SSC 5-Item	0.473	0.588	0.744	0.748	0.604	0.753	0.884	0.891	0.717	0.934	0.958	0.969
0.35	ICT 3-Item	0.279	0.401	0.586	0.628	0.406	0.709	0.799	0.866	0.590	0.847	0.915	0.955
	ICT 4-Item	0.238	0.398	0.537	0.631	0.360	0.619	0.748	0.840	0.582	0.789	0.938	0.930
	ICT 5-Item	0.218	0.361	0.499	0.636	0.299	0.556	0.735	0.807	0.502	0.805	0.871	0.923
	DICT 3-Item	1.013	1.037	0.984	1.016	1.049	0.986	1.004	1.004	1.018	1.025	1.002	0.998
	DICT 4-Item	0.912	0.917	0.999	0.954	0.928	0.970	0.967	0.987	0.994	0.982	0.994	1.006
	DICT 5-Item	0.814	0.949	0.962	0.991	0.907	0.948	0.973	1.001	0.953	0.999	1.005	0.994
	SSC 3-Item	0.702	0.779	0.873	0.913	0.739	0.861	0.919	0.982	0.887	0.981	0.979	0.980
	SSC 4-Item	0.697	0.784	0.906	0.961	0.808	0.898	0.983	0.990	0.912	0.942	0.992	0.982
	SSC 5-Item	0.516	0.699	0.808	0.856	0.626	0.822	0.894	0.956	0.817	0.947	0.974	0.986

Table 12 (continued)

π_s	Non-random Response Model	n = 150				n = 500				n = 1,500			
		Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
0.45	ICT 3-Item	0.312	0.525	0.685	0.740	0.463	0.706	0.853	0.916	0.683	0.908	0.937	0.966
	ICT 4-Item	0.268	0.477	0.659	0.775	0.474	0.738	0.835	0.914	0.675	0.870	0.943	0.983
	ICT 5-Item	0.266	0.426	0.609	0.677	0.379	0.643	0.820	0.862	0.587	0.848	0.941	0.978
	DICT 3-Item	0.972	0.983	0.984	1.011	0.984	0.988	0.969	0.986	0.989	0.981	0.999	1.010
	DICT 4-Item	0.936	0.995	0.960	0.969	0.918	0.974	0.983	1.002	0.950	0.998	0.997	1.005
	DICT 5-Item	0.926	0.942	0.956	0.964	0.944	1.006	0.972	0.997	0.981	0.968	0.991	1.007
	SSC 3-Item	0.768	0.839	0.925	0.976	0.842	0.932	0.970	0.983	0.921	0.975	0.983	0.984
	SSC 4-Item	0.687	0.845	0.917	0.953	0.864	0.945	0.982	0.994	0.877	0.973	1.005	0.990
	SSC 5-Item	0.599	0.744	0.934	0.912	0.742	0.861	0.919	0.953	0.822	0.956	0.979	0.981

*The SSC assumption of using the normal approximation of a binomial for $\boldsymbol{\pi_s}$ =.01 was not met.

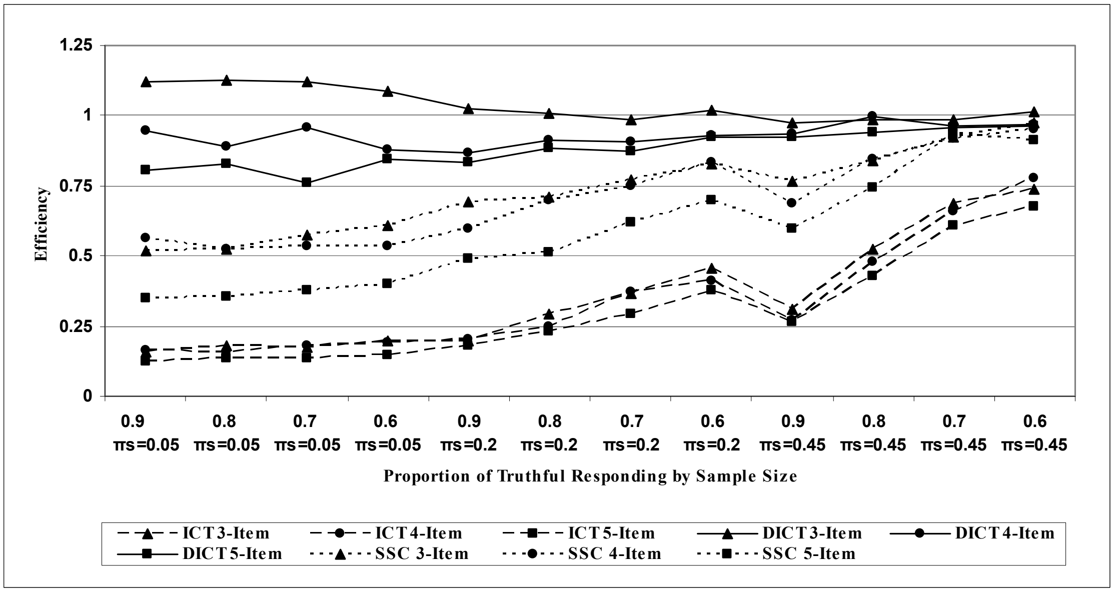


Figure 19. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response Models to the Forced Choice Technique, n = 150

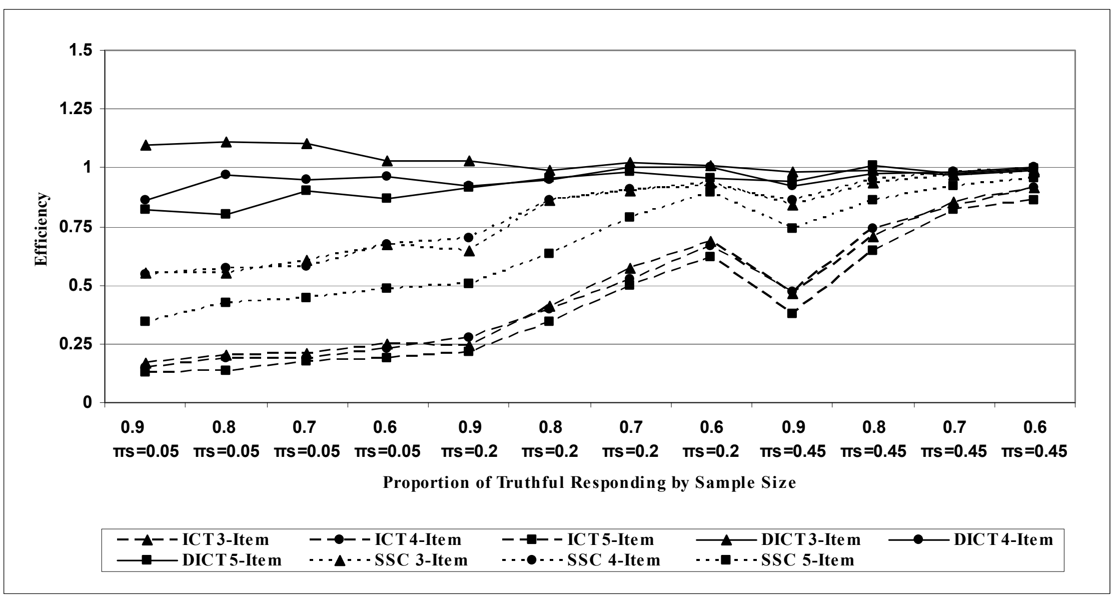


Figure 20. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response Models to the Forced Choice Technique, n = 500

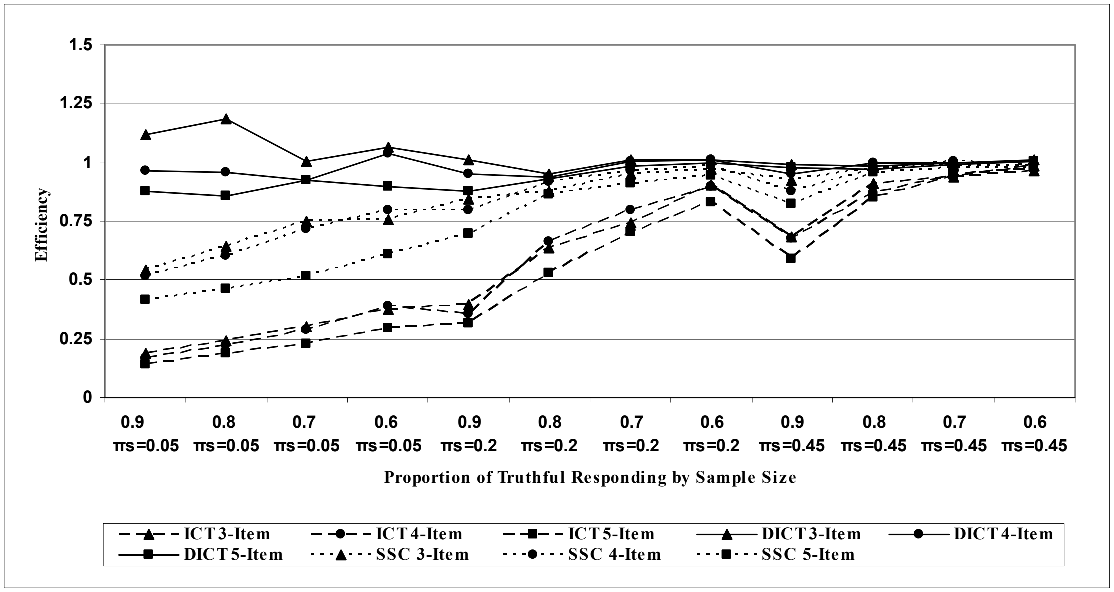


Figure 21. Efficiency as It Relates to Non-compliance, Comparing Optimal Non-random Response models to the Unrelated Question Technique, *n* = 1,500

When π_s was larger (i.e., 0.25, 0.35, and 0.45), as π_s, sample sizes and non-compliance rates increased, the double item count technique (DICT) optimal models continued to be as efficient as the FCT. For optimal single sample count technique (SSC), all models proved, generally to be as efficient as the FCT for increasing π_s, across all sample sizes and non-compliance rates with the exception of models with small sample sizes (i.e., *n* = 150) in the presence of low to moderate non-compliance rates (i.e., percent of truthful reporting $\geq$ 0.80). For the item count technique (ICT) optimal, with the exception of models with large sample sizes (*n* = 1,500) in the presence of moderate to high non-compliance (i.e., percent of truthful reporting $\leq$ 0.70), models were generally *less* efficient than the FCT.

As is evident, in the case of comparing the forced choice technique (FCT) to the non-random response (NRR) optimal models, a pattern was again apparent. Across all sample sizes, sensitive prevalent rates and equivalent non-compliance rates, the double item count technique (DICT) optimal models were most efficient of the NRR optimal models since the ratio of MSE as compared with the FCT was greater than any of the corresponding ratio of MSE of the single sample count technique (SSC) and item count technique (ICT) optimal models. As is seen in each figure, the DICT optimal 3-item list was just as efficient as the FCT, followed by the DICT optimal 4-item and 5-Item list models. Only in the presence of high non-compliance (i.e., percent of truthful reporting $\leq$ 0.70), larger sample sizes (n = 1,500) and larger sensitive prevalent rates (π_s =0.45) did the NRR optimal models and FCT approach unity.

Unrelated Question Technique Optimal vs. Double Item Count Technique Optimal in the Presence of Differing Non-compliant Rates

Because the results of this study indicated that in the presence of equally proportional non-compliance, the unrelated question technique (UQT) optimal was more efficient than all NRR optimal models, a secondary analysis was made following the study of Greenberg et al. (1969). In their study, the authors explored efficiency between Warner's RR technique and the UQT assuming that the UQT improved compliance. Thus, Greenberg et al. (1969) performed an efficiency study comparing the MSE between the two models in cases where compliance was equal or where compliance was improved by the UQT. The authors were then able to conclude that in cases where the UQT *improved* compliance, it was more efficient than Warner's technique (Greenberg et al., 1969). Using this same logic, a secondary analysis exploring efficiency in the presence of differing compliance rates between the UQT optimal and the double item count technique

(DICT) optimal (all list sizes) was made. The DICT optimal models were selected for this analysis since, according to this study, they are the most efficient NRR optimal models. For this analysis, efficiency was compared by selected sensitive prevalent rate meant to represent small ($\pi_s = 0.05$), medium ($\pi_s = 0.20$), and large ($\pi_s = 0.45$) by sample size ($n = 150$, $n = 500$, and $n = 1{,}500$). In examining efficiency, the same principals from the previous analyses were used where models were termed similar if the ratio of the MSE approached unity.

Table 13 displays the results of this analysis. In the table, bolded ratios of MSE indicate cases where the DICT optimal model proved more, or as efficient, as the UQT optimal model. As is evident in the table, when π_s was small ($\pi_s = 0.05$), if the sample size was large ($n = 1{,}500$) and the expected proportion of non-complying in the unrelated question technique (UQT) optimal is higher (i.e., percent of truthful reporting is ≤ 0.70) than the non-compliance in the double item count technique (DICT), than all DICT optimal models were as efficient or more efficient then the UQT optimal. For moderate ($\pi_s = 0.20$) and large ($\pi_s = 0.45$) sensitive prevalent rates, the results were similar. Across all sample sizes, where the percent of truthful reporting in the UQT optimal was *less* than the percent of truthful reporting in all DICT optimal models, the DICT optimal proved to be as efficient or more efficient compared to the UQT optimal. In the case where 90% of respondents of the DICT optimal reported truthfully and only 60% of respondents in the UQT optimal reported truthfully, for moderate and large π_s (i.e., $\pi_s = 0.20$, and $\pi_s = 0.45$) across all sample sizes ($n = 150$, $n = 500$, and $n = 1{,}500$), all DICT optimal models (list sizes 3, 4, and 5) proved to be much more efficient (MSE > 2) than the UQT optimal. As a result, when the DICT optimal models improved compliance substantially over the

UQT optimal, in cases where sensitive prevalent rates were moderate ($\pi_s = 0.20$) or large ($\pi_s = 0.45$), estimates from these models were more efficient. This is also seen with smaller sensitive prevalent rates (π_s=0.05) but only when sample sizes were large (n = 1,500).

The Study of Item-list Size in Non-random Response Models

Efficiency Study Between List Sizes

Tables 14, 15, and 16 displays the results of the efficiency study comparing NRR optimal models (ICT, DICT, and SSC) by their list size. In this analysis, the ratio of MSE within each NRR optimal model was compared by list size. Thus, for each prevalent rate, sample size and percent of truthful reporting (i.e., 100%, 90%, 80%, 70%, and 60%) combination, statistical efficiency was examined by list sizes (i.e., 3-item vs. 4-item, 3-item vs. 5-item, and 4-item vs. 5-item). Comparisons between models were similarly explored as was done in the previous section, where models were termed “similar” if the ratio approached unity. For this analysis, the ratio of MSE was taken in terms of ‘smaller list size’ to ‘larger list size’:

$$Ratio(MSE) = \frac{MSE_{size3}}{MSE_{size4}} \tag{4.1}$$

$$Ratio(MSE) = \frac{MSE_{size3}}{MSE_{size5}} \tag{4.2}$$

$$Ratio(MSE) = \frac{MSE_{size4}}{MSE_{size5}} \tag{4.3}$$

Table 13

Unrelated Question Technique Optimal vs. Double Item Count Technique Optimal Models in the Presence of Differing Non-compliance Rates

Double Item Count Technique		Percent of Truthful Reporting	Unrelated Question Technique Optimal: Percent of Truthful Reporting											
			0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
			$\pi s = 0.05$				$\pi s = 0.20$				$\pi s = 0.45$			
3-Item	$n = 150$	0.9	0.326	0.346	0.396	0.474	0.647	**1.031**	**1.695**	**2.639**	0.843	**2.007**	**3.956**	**6.690**
		0.8	0.318	0.338	0.387	0.463	0.460	0.732	**1.204**	**1.875**	0.385	0.918	**1.808**	**3.058**
		0.7	0.303	0.322	0.368	0.441	0.303	0.483	0.794	**1.237**	0.202	0.480	0.946	**1.600**
		0.6	0.273	0.290	0.332	0.397	0.215	0.343	0.564	0.879	0.124	0.296	0.583	0.985
	$n = 500$	0.9	0.341	0.458	0.666	0.966	0.745	**1.793**	**3.558**	**6.039**	0.917	**2.960**	**6.369**	**11.145**
		0.8	0.310	0.416	0.605	0.878	0.352	0.846	**1.679**	**2.851**	0.299	0.964	**2.073**	**3.628**
		0.7	0.262	0.351	0.511	0.741	0.197	0.475	0.942	**1.599**	0.138	0.445	0.957	**1.674**
		0.6	0.201	0.270	0.393	0.569	0.118	0.284	0.564	0.957	0.081	0.260	0.559	0.978
	$n = 1{,}500$	0.9	0.409	0.779	**1.410**	**2.302**	0.843	**2.689**	**5.777**	**10.106**	0.961	**3.546**	**7.856**	**13.890**
		0.8	0.323	0.615	**1.113**	**1.817**	0.280	0.892	**1.916**	**3.352**	0.264	0.972	**2.155**	**3.810**
		0.7	0.193	0.367	0.665	**1.085**	0.143	0.457	0.982	**1.718**	0.122	0.449	0.994	**1.758**
		0.6	0.144	0.274	0.496	0.809	0.083	0.265	0.569	0.995	0.070	0.257	0.570	**1.007**
4-Item	$n = 150$	0.9	0.274	0.291	0.333	0.398	0.548	0.873	**1.436**	**2.236**	0.812	**1.934**	**3.812**	**6.446**
		0.8	0.252	0.268	0.306	0.367	0.416	0.662	**1.089**	**1.696**	0.390	0.928	**1.830**	**3.094**
		0.7	0.259	0.275	0.314	0.376	0.280	0.446	0.733	**1.141**	0.197	0.468	0.923	**1.561**
		0.6	0.221	0.235	0.269	0.322	0.196	0.312	0.514	0.800	0.119	0.284	0.559	0.945

Table 13 (continued)

Double Item Count Technique		Percent of Truthful Reporting	Unrelated Question Technique Optimal: Percent of Truthful Reporting											
			0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
			πs = 0.05				πs = 0.20				πs = 0.45			
	n = 500	0.9	0.269	0.360	0.524	0.760	0.664	**1.598**	**3.171**	**5.383**	0.856	**2.762**	**5.942**	**10.398**
		0.8	0.270	0.363	0.527	0.765	0.339	0.815	**1.617**	**2.744**	0.294	0.950	2.044	**3.576**
		0.7	0.226	0.303	0.440	0.639	0.193	0.464	0.920	**1.562**	0.140	0.451	0.971	**1.698**
		0.6	0.188	0.252	0.367	0.532	0.117	0.282	0.559	0.948	0.082	0.264	0.568	0.994
	n = 1,500	0.9	0.352	0.672	**1.216**	**1.985**	0.795	**2.535**	**5.445**	**9.526**	0.923	**3.405**	**7.544**	**13.340**
		0.8	0.262	0.499	0.904	**1.475**	0.277	0.884	**1.899**	**3.321**	0.268	0.989	**2.191**	**3.874**
		0.7	0.177	0.338	0.611	0.998	0.142	0.452	0.972	**1.700**	0.121	0.448	0.992	**1.754**
		0.6	0.140	0.266	0.482	0.786	0.083	0.265	0.569	0.996	0.069	0.256	0.567	**1.002**
5-Item	*n* = 150	0.9	0.234	0.249	0.285	0.341	0.527	0.840	**1.381**	**2.151**	0.804	**1.913**	**3.770**	**6.376**
		0.8	0.234	0.248	0.284	0.340	0.403	0.641	**1.055**	**1.642**	0.369	0.879	**1.733**	**2.930**
		0.7	0.205	0.218	0.249	0.298	0.269	0.429	0.705	**1.098**	0.196	0.466	0.919	**1.554**
		0.6	0.213	0.226	0.259	0.310	0.196	0.312	0.512	**0.798**	0.118	0.282	0.555	0.939
	n = 500	0.9	0.256	0.343	0.498	0.723	0.661	**1.590**	**3.155**	**5.356**	0.880	**2.839**	**6.108**	**10.688**
		0.8	0.223	0.300	0.436	0.632	0.340	0.818	**1.623**	**2.755**	0.304	0.981	**2.112**	**3.695**
		0.7	0.215	0.288	0.419	0.608	0.189	0.454	0.900	**1.529**	0.138	0.446	0.960	**1.679**
		0.6	0.170	0.227	0.331	0.480	0.112	0.270	0.535	0.908	0.081	0.263	0.565	0.989

Table 13 (continued)

Double Item Count Technique	Percent of Truthful Reporting	Unrelated Question Technique Optimal: Percent of Truthful Reporting											
		0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
		πs = 0.05				πs = 0.20				πs = 0.45			
n = 1,500	0.9	0.320	0.611	**1.105**	**1.804**	0.735	**2.343**	**5.033**	**8.805**	0.953	**3.517**	**7.792**	**13.778**
	0.8	0.233	0.445	0.805	**1.314**	0.275	0.877	**1.884**	**3.297**	0.260	0.959	**2.125**	**3.758**
	0.7	0.176	0.336	0.608	0.993	0.140	0.446	0.958	**1.675**	0.121	0.445	0.987	**1.745**
	0.6	0.120	0.230	0.415	0.678	0.082	0.261	0.560	0.979	0.069	0.256	0.568	**1.004**

Table 14

Study of Statistics Efficiency by Size of List: Item Count Technique Optimal

π_s	*n*	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.01	150	3-Item	0.800	0.944	0.863	0.871	0.909	0.701	0.656	0.685	0.647	0.769
		4-Item						0.875	0.695	0.794	0.742	0.846
	500	3-Item	1.020	0.878	0.907	0.899	0.970	0.817	0.772	0.756	0.748	0.727
		4-Item						0.802	0.879	0.833	0.832	0.749
	1,500	3-Item	0.956	0.958	0.955	0.899	0.891	0.742	0.723	0.747	0.702	0.703
		4-Item						0.776	0.755	0.782	0.781	0.790
0.03	150	3-Item	0.893	0.903	0.899	0.972	1.023	0.702	0.742	0.697	0.752	0.822
		4-Item						0.787	0.822	0.775	0.773	0.803
	500	3-Item	0.914	0.946	0.890	0.866	0.882	0.676	0.762	0.721	0.762	0.716
		4-Item						0.739	0.806	0.810	0.880	0.811
	1,500	3-Item	0.938	0.876	0.888	1.024	0.901	0.753	0.689	0.697	0.813	0.802
		4-Item						0.803	0.787	0.785	0.794	0.891
0.05	150	3-Item	0.813	1.048	0.882	1.027	0.969	0.740	0.801	0.754	0.771	0.736
		4-Item						0.910	0.764	0.855	0.751	0.760
	500	3-Item	0.926	0.874	0.955	0.891	0.903	0.682	0.744	0.671	0.828	0.747
		4-Item						0.736	0.851	0.703	0.929	0.827
	1,500	3-Item	1.089	0.878	0.918	0.968	1.036	0.840	0.733	0.761	0.769	0.793
		4-Item						0.772	0.836	0.829	0.794	0.765

Table 14 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.10	150	3-Item	0.973	0.973	0.929	0.952	0.924	0.782	0.829	0.759	0.728	0.785
		4-Item						0.804	0.852	0.818	0.764	0.849
	500	3-Item	0.915	0.949	0.984	0.988	0.901	0.775	0.704	0.773	0.864	0.837
		4-Item						0.846	0.741	0.786	0.875	0.929
	1,500	3-Item	0.847	0.995	0.895	1.080	0.958	0.717	0.783	0.741	0.853	0.834
		4-Item						0.846	0.787	0.828	0.790	0.870
0.15	150	3-Item	0.882	1.057	1.002	0.987	0.864	0.809	0.861	0.773	0.773	0.685
		4-Item						0.917	0.814	0.772	0.783	0.792
	500	3-Item	0.941	0.974	0.921	1.051	1.031	0.713	0.823	0.805	0.908	0.828
		4-Item						0.758	0.845	0.874	0.864	0.803
	1,500	3-Item	0.843	0.918	0.898	0.944	0.924	0.774	0.766	0.843	0.884	0.894
		4-Item						0.918	0.834	0.939	0.936	0.967
0.20	150	3-Item	0.843	1.021	0.853	1.012	0.899	0.665	0.906	0.789	0.796	0.822
		4-Item						0.789	0.887	0.925	0.786	0.914
	500	3-Item	0.806	1.148	0.971	0.925	0.973	0.699	0.890	0.831	0.871	0.900
		4-Item						0.867	0.775	0.855	0.942	0.925
	1,500	3-Item	0.859	0.899	1.043	1.076	0.991	0.720	0.797	0.840	0.950	0.920
		4-Item						0.838	0.886	0.805	0.883	0.928

Table 14 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.25	150	3-Item	0.942	0.957	1.047	0.917	0.984	0.821	0.766	0.923	0.850	0.879
		4-Item						0.871	0.801	0.882	0.927	0.893
	500	3-Item	0.844	0.964	0.931	1.052	1.045	0.746	0.843	0.891	0.944	0.867
		4-Item						0.883	0.874	0.957	0.898	0.830
	1,500	3-Item	0.912	0.993	1.045	0.945	1.040	0.760	0.891	0.980	0.936	1.016
		4-Item						0.833	0.898	0.938	0.990	0.977
0.35	150	3-Item	0.898	0.855	0.991	0.917	1.005	0.761	0.783	0.900	0.852	1.014
		4-Item						0.848	0.916	0.908	0.929	1.009
	500	3-Item	0.955	0.886	0.873	0.935	0.970	0.755	0.737	0.784	0.920	0.931
		4-Item						0.791	0.832	0.898	0.983	0.960
	1,500	3-Item	0.978	0.987	0.932	1.026	0.974	0.741	0.851	0.951	0.952	0.967
		4-Item						0.758	0.862	1.020	0.928	0.993
0.45	150	3-Item	1.019	0.861	0.907	0.962	1.047	0.816	0.854	0.811	0.890	0.914
		4-Item						0.800	0.992	0.894	0.925	0.873
	500	3-Item	0.909	1.024	1.045	0.979	0.997	0.776	0.818	0.910	0.961	0.941
		4-Item						0.854	0.799	0.871	0.982	0.943
	1,500	3-Item	0.999	0.989	0.958	1.006	1.018	0.762	0.860	0.934	1.005	1.013
		4-Item						0.762	0.869	0.975	0.999	0.995

Table 15

Study of Statistical Efficiency by Size of List: Double Item Count Technique Optimal

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.01	150	3-Item	0.821	0.841	0.789	0.800	0.857	0.734	0.661	0.684	0.675	0.676
		4-Item						0.894	0.786	0.867	0.844	0.788
	500	3-Item	0.778	0.823	0.846	0.836	0.827	0.665	0.697	0.743	0.798	0.665
		4-Item						0.855	0.848	0.878	0.954	0.804
	1,500	3-Item	0.820	0.830	0.855	0.781	0.877	0.743	0.663	0.741	0.732	0.899
		4-Item						0.906	0.799	0.866	0.937	1.025
0.03	150	3-Item	0.829	0.762	0.804	0.866	0.895	0.756	0.730	0.701	0.756	0.800
		4-Item						0.912	0.958	0.872	0.874	0.893
	500	3-Item	0.715	0.788	0.803	0.777	0.742	0.665	0.758	0.699	0.791	0.701
		4-Item						0.930	0.962	0.870	1.018	0.945
	1,500	3-Item	0.820	0.849	0.945	0.836	0.752	0.720	0.729	0.824	0.725	0.716
		4-Item						0.878	0.859	0.872	0.867	0.952
0.05	150	3-Item	0.838	0.841	0.792	0.853	0.810	0.758	0.719	0.734	0.677	0.780
		4-Item						0.905	0.855	0.926	0.794	0.963
	500	3-Item	0.831	0.787	0.872	0.862	0.935	0.789	0.749	0.720	0.820	0.843
		4-Item						0.949	0.951	0.826	0.951	0.901
	1,500	3-Item	0.803	0.862	0.812	0.920	0.972	0.727	0.784	0.723	0.915	0.838
		4-Item						0.906	0.909	0.891	0.995	0.862

Table 15 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.10	150	3-Item	0.755	0.885	0.897	0.959	0.879	0.738	0.758	0.806	0.795	0.803
		4-Item						0.978	0.857	0.899	0.829	0.914
	500	3-Item	0.883	0.779	1.022	0.950	0.961	0.724	0.711	0.921	0.912	0.906
		4-Item						0.820	0.912	0.901	0.959	0.943
	1,500	3-Item	0.763	0.943	0.880	0.963	0.952	0.662	0.869	0.908	0.935	0.959
		4-Item						0.867	0.922	1.031	0.970	1.007
0.15	150	3-Item	0.861	0.879	0.914	0.875	0.902	0.706	0.777	0.813	0.825	0.823
		4-Item						0.821	0.884	0.890	0.944	0.912
	500	3-Item	0.804	0.978	0.940	0.948	0.975	0.748	0.833	0.837	0.932	0.910
		4-Item						0.931	0.852	0.891	0.983	0.933
	1,500	3-Item	0.889	0.867	0.971	1.022	0.989	0.836	0.843	0.942	0.975	0.989
		4-Item						0.941	0.972	0.970	0.954	1.000
0.20	150	3-Item	0.884	0.847	0.905	0.923	0.910	0.820	0.815	0.876	0.888	0.908
		4-Item						0.928	0.962	0.968	0.962	0.998
	500	3-Item	0.757	0.891	0.963	0.977	0.991	0.825	0.887	0.966	0.956	0.949
		4-Item						1.090	0.995	1.004	0.979	0.957
	1,500	3-Item	1.002	0.943	0.991	0.989	1.001	0.825	0.871	0.983	0.975	0.984
		4-Item						0.823	0.924	0.992	0.985	0.983

Table 15 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.25	150	3-Item	0.871	0.894	0.879	0.919	0.957	0.839	0.808	0.831	0.935	0.998
		4-Item						0.963	0.904	0.945	1.017	1.043
	500	3-Item	0.943	0.922	0.971	1.006	0.993	0.890	0.934	0.952	1.034	0.984
		4-Item						0.944	1.012	0.981	1.028	0.991
	1,500	3-Item	0.866	0.963	0.988	0.990	0.986	0.802	1.008	0.967	0.984	0.992
		4-Item						0.927	1.047	0.979	0.994	1.006
0.35	150	3-Item	0.788	0.901	0.884	1.015	0.939	0.799	0.804	0.915	0.977	0.976
		4-Item						1.014	0.893	1.035	0.963	1.039
	500	3-Item	0.854	0.885	0.984	0.963	0.983	0.856	0.864	0.961	0.969	0.998
		4-Item						1.002	0.977	0.977	1.006	1.014
	1,500	3-Item	0.913	0.977	0.958	0.992	1.008	0.923	0.936	0.975	1.002	0.996
		4-Item						1.011	0.959	1.018	1.010	0.988
0.45	150	3-Item	0.940	0.964	1.012	0.976	0.959	0.815	0.953	0.958	0.971	0.953
		4-Item						0.867	0.989	0.947	0.996	0.994
	500	3-Item	0.977	0.933	0.986	1.015	1.016	0.822	0.959	1.019	1.003	1.010
		4-Item						0.841	1.028	1.033	0.989	0.994
	1,500	3-Item	0.960	0.960	1.017	0.998	0.995	0.874	0.992	0.986	0.993	0.997
		4-Item						0.911	1.033	0.970	0.995	1.002

Table 16

Study of Statistical Efficiency by Size of List: Single Sample Count Optimal

π_s	*n*	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.03	150	3-Item	0.892	1.008	0.955	0.950	1.021	0.675	0.634	0.610	0.683	0.625
		4-Item						0.757	0.629	0.639	0.719	0.613
	500	3-Item	0.993	1.213	0.956	0.982	1.041	0.615	0.794	0.642	0.617	0.695
		4-Item						0.619	0.654	0.672	0.629	0.667
	1,500	3-Item	1.039	0.956	0.962	1.004	1.025	0.634	0.646	0.685	0.717	0.776
		4-Item						0.610	0.675	0.712	0.714	0.758
0.05	150	3-Item	1.113	1.088	1.006	0.931	0.877	0.734	0.671	0.678	0.654	0.661
		4-Item						0.660	0.617	0.674	0.702	0.753
	500	3-Item	1.002	0.997	1.029	0.955	1.003	0.709	0.628	0.769	0.735	0.724
		4-Item						0.707	0.630	0.747	0.770	0.722
	1,500	3-Item	1.111	0.953	0.936	0.956	1.049	0.662	0.761	0.720	0.685	0.802
		4-Item						0.596	0.799	0.769	0.717	0.765

Table 16 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.10	150	3-Item	1.021	1.068	0.908	0.917	0.988	0.678	0.751	0.683	0.784	0.682
		4-Item						0.664	0.703	0.752	0.854	0.691
	500	3-Item	1.078	1.011	0.966	0.979	0.996	0.736	0.699	0.743	0.788	0.905
		4-Item						0.683	0.692	0.769	0.805	0.909
	1,500	3-Item	1.148	1.044	1.002	0.985	0.939	0.792	0.725	0.822	0.825	0.915
		4-Item						0.690	0.694	0.820	0.837	0.974
0.15	150	3-Item	0.971	0.980	0.919	1.197	1.032	0.700	0.765	0.723	0.828	0.817
		4-Item						0.721	0.781	0.787	0.692	0.792
	500	3-Item	0.960	0.972	0.992	0.945	0.998	0.690	0.758	0.786	0.848	0.950
		4-Item						0.719	0.780	0.792	0.897	0.952
	1,500	3-Item	1.036	1.067	1.017	0.977	1.016	0.762	0.812	0.857	0.953	0.958
		4-Item						0.735	0.761	0.843	0.975	0.943
0.20	150	3-Item	0.966	0.863	0.982	0.975	1.005	0.742	0.707	0.722	0.804	0.843
		4-Item						0.769	0.819	0.735	0.825	0.839
	500	3-Item	0.867	1.085	0.997	1.008	0.983	0.600	0.781	0.737	0.874	0.957
		4-Item						0.691	0.720	0.739	0.867	0.974
	1,500	3-Item	1.036	0.946	1.045	1.014	1.014	0.780	0.828	0.981	0.956	0.968
		4-Item						0.753	0.875	0.938	0.943	0.955

Table 16 (continued)

π_s	n	List Size	Percent of Truthful Reporting									
			4-Item					5-Item				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
0.25	150	3-Item	0.951	0.945	1.187	0.957	0.937	0.733	0.736	0.906	0.898	0.816
		4-Item						0.771	0.779	0.763	0.938	0.871
	500	3-Item	0.971	0.892	1.085	1.002	0.971	0.730	0.803	0.930	0.968	0.946
		4-Item						0.752	0.900	0.857	0.966	0.974
	1,500	3-Item	0.942	1.018	0.964	1.019	1.012	0.723	0.882	0.989	1.004	0.991
		4-Item						0.767	0.866	1.026	0.985	0.979
0.35	150	3-Item	1.027	0.993	1.007	1.038	1.053	0.736	0.735	0.898	0.926	0.937
		4-Item						0.716	0.740	0.891	0.893	0.890
	500	3-Item	0.981	1.094	1.043	1.070	1.008	0.727	0.847	0.955	0.973	0.974
		4-Item						0.742	0.774	0.915	0.910	0.966
	1,500	3-Item	1.014	1.028	0.961	1.013	1.002	0.696	0.921	0.965	0.995	1.006
		4-Item						0.686	0.896	1.005	0.983	1.004
0.45	150	3-Item	1.012	0.895	1.007	0.991	0.976	0.741	0.780	0.887	1.010	0.934
		4-Item						0.732	0.872	0.881	1.019	0.957
	500	3-Item	0.993	1.026	1.014	1.012	1.011	0.762	0.881	0.923	0.947	0.969
		4-Item						0.768	0.859	0.910	0.936	0.958
	1,500	3-Item	1.092	0.953	0.999	1.022	1.006	0.778	0.893	0.981	0.996	0.997
		4-Item						0.713	0.937	0.983	0.975	0.990

In examining the results of the analysis, it became evident that across all NRR optimal models, list size 3 models were nearly as efficient as list size 4-models (see Tables 14, 15, 16). The more important analysis was in examining the efficiency between 3-item and 5-item models, since it was assumed that NRR models with larger list sizes provided higher levels of confidentiality which in theory resulted in a higher percentage of truthful reporting. As is evident in each table, the 3-item models were generally more efficient (i.e., MSE < 1) compared with the 5-item models when participants report truthfully (i.e., 100% truthful reporting). As truthful reporting declined from 100% to 60%, a trend in model efficiency was apparent for each NRR optimal model where the 3-item and 5-item lists became more similar in terms of efficiency and in fact approached unity. For each table, increasing efficiency trends are bolded. The difference in these trends between each NRR optimal model was where the trend begins and when unity was reached. For the optimal item count technique (ICT; Table 14), evidence of a trend in increasing efficiency between the 3-item and 5-item optimal model was most apparent when sensitive prevalent rates were large, (i.e., $\pi_s = 0.35$, $\pi_s = 0.45$) across all sample sizes ($n = 150$, $n = 500$, and $n = 1{,}500$) where unity was reached when non-compliance rates in both models was highest (i.e., 60% truthful reporting). For the optimal double item count technique (DICT; Table 15), evidence of increasing efficiency between the 3-item and 5-item model in the presence of increasing non-compliance appeared when sensitive prevalent rates were moderate (i.e., $\pi_s = 0.10$) across all sample sizes where unity was approached when non-compliance in both models was highest (i.e., 60% truthful reporting). Likewise, for the optimal SSC, evidence of increasing efficiency between the 3-item and 5-item model in the presence of increasing non-compliance also appeared when sensitive prevalent rates were moderate. For larger sample sizes, (n = 500, n =1,500), the trend was evident at $\pi_s = 0.10$, whereas for smaller sample sizes the trend began at $\pi_s = 0.20$ where unity was approached when non-compliance between both models was highest (i.e., 60% truthful reporting).

In addition, for the optimal item count technique (ICT) and double item count technique (DICT) models, 3-item and 5-item models became more similar, as non-compliance increased, when sensitive prevalent rates were moderate (i.e., $\pi_s = 0.25$ for ICT optimal, $\pi_s = 0.20$ for DICT optimal) across all sample sizes (n = 150, n = 500, and n = 1,500) whereas for the SSC optimal, 3-item, and 5-item models became more similar for large sensitive prevalent rates ($\pi_s = 0.45$) across all sample sizes.

Item List 5 vs. Item List 3: By Optimal Non-random Response Model in the Presence of Differing Non-compliant Rates

Because the results of this study indicated that in the presence of equally proportional non-compliance, generally NRR optimal models with 3-item lists were more efficient compared to models with 5-item lists, a secondary analysis exploring the effects on efficiency between the two models in cases where compliance are equal or improved by the 5-item list were made. This analysis is similar to the secondary analysis that explored efficiency between the optimal DICT and optimal unrelated question technique (UQT) in the presence of differing compliance rates. As was done in the previous analysis, efficiency was compared by selected sensitive prevalent rate meant to represent small ($\pi_s = 0.05$), medium ($\pi_s = 0.20$), and large ($\pi_s = 0.45$) by sample size (n = 150, n = 500, and n = 1,500) and NRR optimal model (DICT, SSC, and ICT). In examining efficiency, the same principals from the previous analyses were used where models were termed similar if the ratio of MSE approached unity.

Table 17 displays the results of this analysis. In the table, bolded ratios of MSE indicated cases where the 5-item model and the 3-item model were greater than or equal to unity (i.e., ratio of MSE $\geq$ 1). As is evident in the table, for the DICT optimal when π_s was small ($\pi_s = 0.05$), if the sample size was also small (n = 150), the 3-item double item count technique (DICT) optimal was generally more efficient than the 5-item DICT optimal. As sample sizes became moderate (n = 500) for π_s small ($\pi_s = 0.05$), when compliant rates decreased substantially in the 3-item DICT (i.e., percent of truthful reporting $\leq$ 0.70) compared to the DICT optimal 5-item model, the 5-item model proved just as efficient. For larger sample sizes (n = 1,500) when π_s was small ($\pi_s = 0.05$), if compliant rates decreased substantially in the 3-item DICT (i.e., percent of truthful reporting $\leq$ 0.70) compared to the DICT optimal 5-item model, the 5-item model was more efficient. For moderate ($\pi_s = 0.20$) and large ($\pi_s = 0.45$) sensitive prevalent rates, results were similar across all sample sizes, where if the optimal DICT 5-item model improved compliant rates compared to the DICT 3-item model, the DICT 5-item proved to be more efficient.

For the single sample count technique (SSC), when the sensitive prevalent rate was small ($\pi_s = 0.05$), unless sample sizes were large (n = 1,500) and the non-compliance rate of the SSC optimal 3-Item model was high (i.e., percent of truthful reporting $\leq$ 0.70) compared to the non-compliance rate of the optimal 5-item model (i.e., percent of truthful reporting $\leq$ 0.90), the 5-item model was more efficient. When the sensitive prevalent rate was moderate ($\pi_s = 0.20$), in smaller sample sizes (n = 150), when the non-compliance rate of the SSC optimal 3-item model was high (i.e., percent of truthful reporting = 0.60) compared to the non-compliance rate of the optimal 5-Item model (i.e., percent of truthful reporting $\geq$ 0.80), the SSC 5-item model was more efficient. For mid (n = 500) to large sample sizes (n = 1,500), when sensitive prevalent rates were moderate ($\pi_s = 0.20$), if the SSC optimal 5-item model improved compliance compared to the optimal 3-item model, the 5-item model was more efficient. For large sensitive prevalent rates ($\pi_s = 0.45$), across all sample sizes, when the single sample count technique (SSC) optimal 5-item model improved the compliance rate compared to the SSC optimal 3-item model, the 5-item model was more efficient.

For the item count technique (ICT), when the sensitive prevalent rate was small ($\pi_s = 0.05$), in general, the ICT optimal 3-item model was more efficient compared with the 5-iItem model across all sample sizes and compliance rates. When the sensitive prevalent rate was moderate ($\pi_s = 0.20$), unless sample sizes were larger (n = 500, n = 1,500) if the ICT 5-Item optimal model improved the compliance rate of the ICT-3 optimal model, the 5-Item model was more efficient. For large sensitive prevalent rates ($\pi_s = 0.45$), when sample sizes were small (n = 150) and the non-compliance rate of the ICT optimal 3-item model was high (i.e., percent of truthful reporting $\leq$ 0.70) compared to the non-compliance rate of the optimal ICT 5-item model (i.e., percent of truthful reporting $\geq$ 0.90), the 5-item model was more efficient. As sample sizes increased, for large sensitive prevalent rates ($\pi_s = 0.45$), if the ICT optimal 5-item model improved the compliance rate compared to the ICT optimal 3-item model, the 5-item model was more efficient (see Table 17)

Table 17

Item List 3 vs. Item List 5: Non-random Response Optimal Models in the Presence of Differing Non-compliance Rates

	List Size	*n*	Percent of Truthful	3-Item: Percent of Truthful Reporting														
				1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
				πs = 0.05					πs = 0.20					πs = 0.45				
Double Item Count Technique	5	150	1.0	0.76	0.70	0.71	0.75	0.83	0.82	0.93	**1.31**	**1.99**	**2.80**	0.82	**1.25**	**2.73**	**5.21**	**8.47**
			0.9	0.78	0.72	0.74	0.77	0.86	0.72	0.81	**1.15**	**1.74**	**2.45**	0.62	0.95	**2.08**	**3.98**	**6.47**
			0.8	0.78	0.72	0.73	0.77	0.86	0.55	0.62	0.88	**1.33**	**1.87**	0.29	0.44	0.96	**1.83**	**2.97**
			0.7	0.69	0.63	0.64	0.68	0.75	0.37	0.42	0.59	0.89	**1.25**	0.15	0.23	0.51	0.97	**1.58**
			0.6	0.71	0.65	0.67	0.70	0.78	0.27	0.30	0.43	0.65	0.91	0.09	0.14	0.31	0.59	0.95
		500	1.0	0.79	0.81	0.90	**1.06**	**1.38**	0.82	**1.29**	**2.74**	**4.88**	**8.16**	0.82	**2.63**	**8.07**	**17.49**	**29.92**
			0.9	0.73	0.75	0.82	0.98	**1.27**	0.57	0.89	**1.88**	**3.35**	**5.60**	0.30	0.96	**2.95**	**6.39**	**10.92**
			0.8	0.63	0.65	0.72	0.85	**1.11**	0.29	0.46	0.97	**1.72**	**2.88**	0.10	0.33	**1.02**	**2.21**	**3.78**
			0.7	0.61	0.63	0.69	0.82	**1.07**	0.16	0.25	0.54	0.96	**1.60**	0.05	0.15	0.46	**1.00**	**1.72**
			0.6	0.48	0.50	0.55	0.65	0.84	0.10	0.15	0.32	0.57	0.95	0.03	0.09	0.27	0.59	**1.01**
		1,500	1.0	0.73	0.85	**1.08**	**1.81**	**2.42**	0.82	**1.97**	**5.93**	**11.56**	**19.96**	0.87	6.55	**23.86**	**51.72**	**90.28**
			0.9	0.67	0.78	0.99	**1.66**	**2.23**	0.37	0.87	**2.63**	**5.12**	**8.85**	0.13	0.99	**3.62**	**7.84**	**13.68**
			0.8	0.49	0.57	0.72	**1.21**	**1.62**	0.14	0.33	0.98	**1.92**	**3.31**	0.04	0.27	0.99	**2.14**	**3.73**
			0.7	0.37	0.43	0.55	0.92	**1.23**	0.07	0.17	0.50	0.97	**1.68**	0.02	0.13	0.46	0.99	**1.73**
			0.6	0.25	0.29	0.37	0.62	0.84	0.04	0.10	0.29	0.57	0.98	0.01	0.07	0.26	0.57	1.00

Table 17 (continued)

	List Size	n	Percent of Truthful	Ite-3: Percent of Truthful Reporting 1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
				πs = 0.05					πs = 0.20					πs = 0.45				
Single Sample Count	5	150	1.0	0.73	0.68	0.69	0.65	0.66	0.74	0.69	0.94	**1.28**	**1.73**	0.74	0.97	**1.97**	**3.42**	**5.40**
			0.9	0.73	0.67	0.68	0.65	0.66	0.76	0.71	0.96	**1.31**	**1.77**	0.59	0.78	**1.58**	**2.74**	**4.33**
			0.8	0.72	0.67	0.68	0.64	0.65	0.57	0.53	0.72	0.99	**1.34**	0.33	0.44	0.89	**1.54**	**2.43**
			0.7	0.74	0.68	0.69	0.65	0.66	0.47	0.43	0.59	0.80	**1.09**	0.22	0.29	0.58	**1.01**	**1.60**
			0.6	0.73	0.68	0.69	0.65	0.66	0.36	0.34	0.46	0.62	0.84	0.13	0.17	0.34	0.59	0.93
		500	1.0	0.71	0.69	0.76	0.81	0.89	0.60	0.91	**1.38**	**2.45**	**3.88**	0.76	**1.74**	**4.86**	**9.92**	**17.05**
			0.9	0.65	0.63	0.69	0.74	0.82	0.52	0.78	**1.19**	**2.11**	**3.34**	0.39	0.88	**2.45**	**5.01**	**8.61**
			0.8	0.72	0.70	0.77	0.82	0.91	0.32	0.48	0.74	**1.31**	**2.07**	0.14	0.33	0.92	1.89	**3.24**
			0.7	0.64	0.62	0.69	0.73	0.81	0.21	0.32	0.49	0.87	**1.38**	0.07	0.17	0.46	0.95	**1.63**
			0.6	0.57	0.56	0.61	0.66	0.72	0.15	0.22	0.34	0.60	0.96	0.04	0.10	0.28	0.56	0.97
		1,500	1.0	0.66	0.71	0.81	0.98	**1.38**	0.78	**1.30**	**3.53**	**6.78**	**11.48**	0.78	**3.69**	**12.60**	**27.55**	**48.57**
			0.9	0.71	0.76	0.86	**1.05**	**1.47**	0.50	0.83	**2.26**	**4.33**	**7.33**	0.19	0.89	**3.05**	**6.67**	**11.77**
			0.8	0.59	0.64	0.72	0.88	**1.23**	0.22	0.36	0.98	**1.88**	**3.19**	0.06	0.29	0.98	**2.15**	**3.78**
			0.7	0.46	0.50	0.56	0.69	0.96	0.11	0.18	0.50	0.96	**1.62**	0.03	0.13	0.46	1.00	**1.76**
			0.6	0.39	0.42	0.47	0.57	0.80	0.07	0.11	0.30	0.57	0.97	0.02	0.08	0.26	0.57	1.00

Table 17 (continued)

	List Size	*n*	Percent of Truthful	Item-3: Percent of Truthful Reporting														
				1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
				πs = 0.05					πs = 0.20					πs = 0.45				
Item Count Technique	5	150	1.0	0.74	0.81	0.73	0.77	0.74	0.66	0.77	0.73	0.85	1.00	0.82	0.89	**1.16**	**1.71**	**2.63**
			0.9	0.73	0.80	0.72	0.76	0.74	0.79	0.91	0.86	**1.01**	**1.18**	0.79	0.85	1.12	**1.64**	**2.54**
			0.8	0.77	0.83	0.75	0.80	0.77	0.72	0.83	0.79	0.92	**1.09**	0.57	0.62	0.81	**1.19**	**1.84**
			0.7	0.74	0.81	0.73	0.77	0.75	0.62	0.72	0.68	0.80	0.94	0.43	0.46	0.61	0.89	**1.37**
			0.6	0.73	0.80	0.72	0.76	0.74	0.55	0.63	0.60	0.70	0.82	0.28	0.31	0.40	0.59	0.91
		500	1.0	0.68	0.71	0.66	0.74	0.76	0.70	0.94	**1.10**	**1.47**	**2.01**	0.78	**1.15**	**2.33**	**4.10**	**6.65**
			0.9	0.71	0.74	0.70	0.78	0.79	0.66	0.89	**1.05**	**1.40**	**1.91**	0.55	0.82	1.65	**2.91**	**4.72**
			0.8	0.69	0.72	0.67	0.75	0.77	0.53	0.70	0.83	**1.11**	**1.52**	0.30	0.45	0.91	**1.60**	**2.60**
			0.7	0.76	0.79	0.74	0.83	0.84	0.41	0.55	0.65	0.87	**1.19**	0.18	0.27	0.55	0.96	**1.56**
			0.6	0.67	0.70	0.65	0.73	0.75	0.31	0.42	0.49	0.66	0.90	0.11	0.16	0.33	0.58	0.94
		1,500	1.0	0.84	0.76	0.78	0.91	**1.03**	0.72	0.97	**1.71**	**3.04**	**4.30**	0.76	**1.85**	**5.04**	**10.78**	**18.46**
			0.9	0.81	0.73	0.75	0.87	0.99	0.59	0.80	**1.41**	**2.51**	**3.55**	0.35	0.86	**2.34**	**5.00**	**8.56**
			0.8	0.82	0.74	0.76	0.88	**1.00**	0.35	0.47	0.84	**1.49**	**2.11**	0.14	0.34	0.93	**2.00**	**3.42**
			0.7	0.71	0.65	0.66	0.77	0.87	0.23	0.30	0.54	0.95	**1.35**	0.07	0.17	0.47	**1.00**	**1.72**
			0.6	0.65	0.59	0.60	0.70	0.79	0.15	0.21	0.37	0.65	0.92	0.04	0.10	0.28	0.59	**1.01**

As was evident in this analysis, NRR 5-item models were generally more efficient than the corresponding 3-item models when compliance was improved by the 5-item model, depending on the NRR model (DICT, ICT, and SSC), sample size, and sensitive prevalent rate. However, what was clear from this analysis, was that the DICT optimal 5-item list models proved to be more efficient compared with the optimal DICT 3-item list models under the assumption that 5-item models improved compliance rates, at lower sensitive prevalent rates ($\pi_S = 0.05$) and smaller sample sizes ($n = 500$) compared to the optimal SSC and optimal ICT models.

Efficiency Study Between Non-random Response Models

Table 18 displays the results of the efficiency study comparing NRR optimal models, by list size (3-item, 4-item, and 5-item), sample size ($n = 150$, $n = 500$, and $n = 1{,}500$) and compliant rate (1, 0.90, 0.80, 0.70, and 0.60) between NRR optimal models (ICT, DICT, and SSC). Comparisons between models were similarly explored as was done in the previous sections, where models were termed "similar" if the ratio of MSE approaches unity. Ratios of MSE for which models are termed similar are bolded in the table. For this analysis, the ratios of MSE were made as follows for each list size comparison:

$$Ratio(MSE) = \frac{MSE_{ICT}}{MSE_{DICT}} \tag{4.4}$$

$$Ratio(MSE) = \frac{MSE_{SSC}}{MSE_{DICT}} \tag{4.5}$$

$$Ratio(MSE) = \frac{MSE_{SSC}}{MSE_{ICT}} \tag{4.6}$$

Table 18

Statistical Efficiency: Comparisons Between Non-random Response Optimal Models by Size of List

List Size	πs	n	Percent of Truthful Reporting														
			Ratio of MSE: ICT/DICT					Ratio of MSE: SSC/DICT					Ratio of MSE: ICT/SSC				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
3-Item	0.01	150	6.878	6.812	6.753	6.677	7.641	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		500	7.755	6.780	6.926	6.894	7.442	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		1,500	7.349	7.497	7.088	6.354	5.789	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	0.03	150	6.749	7.009	6.862	7.021	6.631	2.134	2.105	1.977	2.062	1.920	3.162	3.330	3.471	3.405	3.453
		500	7.493	6.721	6.557	6.380	6.020	2.257	2.415	1.982	2.033	2.116	3.320	2.784	3.308	3.138	2.845
		1,500	7.054	6.385	5.483	5.345	4.725	2.083	2.013	1.765	1.723	1.832	3.387	3.172	3.107	3.103	2.580
	0.05	150	6.031	7.161	6.328	6.357	5.536	2.162	2.170	2.158	1.948	1.779	2.790	3.300	2.932	3.264	3.111
		500	6.459	6.522	5.541	5.243	4.100	2.133	1.998	2.007	1.815	1.535	3.029	3.265	2.760	2.889	2.672
		1,500	7.732	5.981	4.855	3.361	2.845	2.247	2.071	1.843	1.343	1.406	3.441	2.888	2.634	2.502	2.024
	0.10	150	6.247	5.872	5.035	4.528	4.190	2.038	2.000	1.651	1.612	1.593	3.066	2.935	3.049	2.810	2.630
		500	6.215	5.869	3.810	3.187	2.500	2.056	1.983	1.440	1.407	1.336	3.023	2.959	2.645	2.266	1.871
		1,500	6.184	4.375	2.591	2.038	1.544	2.212	1.542	1.400	1.166	1.090	2.796	2.838	1.850	1.748	1.417
	0.15	150	5.692	5.661	4.599	3.614	2.547	1.890	1.713	1.601	1.616	1.366	3.012	3.304	2.873	2.236	1.865
		500	5.951	4.371	2.835	2.279	1.806	1.910	1.579	1.354	**1.229**	**1.171**	3.116	2.768	2.094	1.854	1.542
		1,500	5.079	3.540	1.840	1.395	1.220	1.720	1.455	**1.193**	**1.071**	**1.081**	2.953	2.432	1.543	**1.302**	**1.129**
	0.20	150	5.086	5.136	3.459	2.671	2.233	1.802	1.476	1.419	1.280	1.230	2.822	3.480	2.437	2.086	1.815
		500	5.044	4.316	2.401	1.795	1.470	1.648	1.595	**1.146**	**1.137**	**1.079**	3.062	2.706	2.096	1.578	1.363
		1,500	4.531	2.551	1.500	1.363	**1.119**	1.711	**1.194**	**1.079**	**1.061**	**1.041**	2.648	2.136	1.390	**1.284**	**1.076**

Table 18 (continued)

List Size	πs	n	Percent of Truthful Reporting														
			Ratio of MSE: ICT/DICT					Ratio of MSE: SSC/DICT					Ratio of MSE: ICT/SSC				
			1	0.9	0.8	0.7	0.6	1	0.9	0.8	0.7	0.6	1	0.9	0.8	0.7	0.6
	0.25	150	5.249	4.533	3.479	2.338	1.793	1.666	1.620	1.604	**1.243**	**1.090**	3.150	2.798	2.169	1.881	1.645
		500	4.758	3.076	1.913	1.471	1.306	1.637	1.263	**1.241**	**1.045**	**1.076**	2.907	2.435	1.542	1.407	1.214
		1,500	4.956	2.080	1.484	**1.126**	**1.142**	1.721	**1.200**	**1.064**	**1.048**	**1.025**	2.879	1.734	1.394	**1.075**	**1.114**
	0.35	150	5.131	3.634	2.584	1.679	1.618	1.722	1.442	1.331	**1.128**	**1.113**	2.980	2.519	1.941	1.489	1.454
		500	4.882	2.586	1.390	**1.256**	**1.158**	1.653	1.420	1.145	**1.093**	**1.022**	2.953	1.820	1.214	**1.149**	**1.133**
		1,500	4.499	1.725	1.210	**1.096**	**1.045**	1.559	1.148	**1.045**	**1.024**	**1.019**	2.885	1.503	**1.158**	**1.070**	**1.026**
	0.45	150	4.397	3.120	1.872	1.437	1.365	1.475	1.266	**1.171**	**1.064**	**1.035**	2.981	2.464	1.598	1.350	**1.318**
		500	4.574	2.127	1.398	**1.135**	**1.076**	1.633	**1.169**	**1.060**	0.999	**1.003**	2.800	1.819	1.320	**1.137**	**1.073**
		1,500	4.457	1.448	**1.080**	**1.066**	**1.046**	1.699	**1.074**	**1.007**	**1.016**	**1.026**	2.624	1.348	**1.073**	**1.049**	**1.019**
4-Item	0.01	150	7.053	6.070	6.178	6.135	7.210	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		500	5.918	6.352	6.465	6.416	6.342	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		1,500	6.306	6.496	6.347	5.526	5.700	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	0.03	150	6.264	5.915	6.134	6.252	5.805	1.984	1.591	1.665	1.880	1.684	3.157	3.717	3.685	3.326	3.448
		500	5.862	5.600	5.919	5.723	5.062	1.625	1.569	1.665	1.609	1.508	3.606	3.569	3.555	3.558	3.356
		1,500	6.171	6.187	5.836	4.368	3.945	1.645	1.786	1.734	1.436	1.344	3.751	3.464	3.366	3.042	2.935
	0.05	150	6.214	5.744	5.686	5.283	4.629	1.627	1.678	1.699	1.784	1.644	3.819	3.423	3.347	2.961	2.816
		500	5.795	5.873	5.060	5.074	4.245	1.769	1.578	1.700	1.639	1.430	3.275	3.722	2.976	3.096	2.969
		1,500	5.705	5.875	4.293	3.192	2.667	1.625	1.874	1.598	1.291	1.303	3.511	3.134	2.686	2.472	2.048

Table 18 (continued)

List Size	πs	n	Percent of Truthful Reporting														
			Ratio of MSE: ICT/DICT					Ratio of MSE: SSC/DICT					Ratio of MSE: ICT/SSC				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
	0.10	150	4.847	5.342	4.862	4.560	3.985	1.506	1.657	1.631	1.685	1.417	3.219	3.224	2.982	2.706	2.812
		500	5.993	4.817	3.959	3.066	2.666	1.684	1.529	1.525	1.365	1.289	3.559	3.151	2.596	2.245	2.068
		1,500	5.578	4.147	2.550	1.818	1.535	1.470	1.391	1.231	1.140	1.105	3.793	2.980	2.072	1.595	1.390
	0.15	150	5.552	4.705	4.195	3.204	2.658	1.676	1.537	1.591	**1.181**	**1.193**	3.313	3.061	2.636	2.712	2.228
		500	5.087	4.390	2.893	2.055	1.709	1.599	1.589	1.283	**1.232**	**1.145**	3.181	2.763	2.256	1.668	1.493
		1,500	5.352	3.342	1.990	1.510	**1.305**	1.476	1.183	**1.139**	**1.120**	**1.052**	3.627	2.825	1.747	1.348	1.241
	0.20	150	5.329	4.263	3.668	2.435	2.260	1.649	1.450	**1.307**	**1.212**	**1.114**	3.231	2.941	2.808	2.009	2.028
		500	4.737	3.351	2.380	1.897	1.497	1.437	1.310	**1.106**	**1.103**	**1.087**	3.296	2.557	2.151	1.720	1.377
		1,500	5.286	2.673	1.425	**1.253**	**1.130**	1.655	1.189	**1.023**	**1.035**	**1.027**	3.193	2.249	1.393	**1.210**	**1.101**
	0.25	150	4.856	4.235	2.921	2.344	1.743	1.526	1.532	**1.188**	**1.193**	**1.113**	3.181	2.765	2.459	1.965	1.566
		500	5.311	2.942	1.994	1.406	1.241	1.588	1.306	**1.110**	**1.049**	**1.100**	3.344	2.253	1.797	1.341	**1.128**
		1,500	4.702	2.016	1.403	**1.180**	**1.082**	1.582	**1.135**	**1.091**	**1.018**	**0.998**	2.972	1.777	1.286	1.159	**1.085**
	0.35	150	4.503	3.830	2.306	1.860	1.512	1.321	**1.308**	**1.169**	**1.104**	**0.993**	3.408	2.928	1.972	1.686	1.523
		500	4.366	2.581	1.567	**1.294**	**1.174**	1.440	**1.148**	**1.080**	**0.984**	**0.997**	3.032	2.247	1.451	**1.314**	**1.178**
		1,500	4.200	1.707	**1.245**	**1.060**	**1.082**	1.403	**1.090**	**1.042**	**1.003**	**1.025**	2.993	1.566	**1.194**	**1.057**	**1.056**
	0.45	150	4.057	3.490	2.087	1.458	1.251	1.372	1.363	**1.177**	**1.047**	**1.018**	2.958	2.560	1.773	**1.392**	**1.229**
		500	4.916	1.938	**1.319**	**1.177**	**1.097**	1.607	**1.063**	**1.030**	**1.001**	**1.008**	3.060	1.823	**1.281**	**1.176**	**1.087**
		1,500	4.279	1.406	**1.147**	**1.057**	**1.022**	1.493	**1.082**	**1.025**	**0.992**	**1.014**	2.866	1.299	**1.119**	**1.066**	**1.008**

Table 18 (continued)

List Size	πs	n	Percent of Truthful Reporting														
			Ratio of MSE: ICT/DICT					Ratio of MSE: SSC/DICT					Ratio of MSE: ICT/SSC				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
5-Item	0.01	150	7.206	6.867	6.746	6.973	6.715	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		500	6.312	6.127	6.809	7.354	6.811	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		1,500	7.360	6.881	7.031	6.629	7.402	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	0.03	150	7.262	6.892	6.903	7.063	6.455	2.389	2.423	2.274	2.283	2.456	3.040	2.844	3.036	3.094	2.628
		500	7.370	6.685	6.355	6.621	5.893	2.441	2.306	2.157	2.604	2.135	3.019	2.899	2.947	2.543	2.761
		1,500	6.742	6.757	6.483	4.769	4.216	2.367	2.274	2.123	1.744	1.689	2.848	2.972	3.054	2.735	2.496
	0.05	150	6.180	6.431	6.157	5.586	5.867	2.231	2.325	2.335	2.017	2.101	2.770	2.766	2.637	2.769	2.793
		500	7.472	6.562	5.949	5.196	4.626	2.373	2.381	1.880	2.026	1.786	3.148	2.756	3.164	2.565	2.590
		1,500	6.692	6.391	4.611	3.999	3.005	2.469	2.131	1.852	1.793	1.468	2.711	2.999	2.489	2.230	2.047
	0.10	150	5.896	5.371	5.346	4.946	4.287	2.219	2.019	1.949	1.634	1.875	2.658	2.660	2.743	3.027	2.287
		500	5.808	5.928	4.538	3.362	2.708	2.023	2.015	1.786	1.626	1.337	2.871	2.942	2.541	2.067	2.024
		1,500	5.714	4.854	3.176	2.232	1.777	1.849	1.847	1.548	1.321	1.142	3.090	2.628	2.052	1.690	1.555
	0.15	150	4.969	5.107	4.833	3.862	3.059	1.907	1.739	1.798	1.611	1.374	2.606	2.937	2.688	2.397	2.226
		500	6.244	4.426	2.947	2.337	1.986	2.070	1.735	1.442	1.351	**1.122**	3.017	2.551	2.044	1.730	1.770
		1,500	5.485	3.894	2.056	1.539	1.350	1.889	1.511	**1.311**	**1.097**	**1.115**	2.904	2.577	1.569	1.404	1.210
	0.20	150	6.268	4.622	3.839	2.982	2.466	1.990	1.702	1.722	**1.415**	**1.325**	3.149	2.716	2.230	2.107	1.861
		500	5.952	4.301	2.792	1.969	1.550	2.265	1.810	1.502	**1.244**	**1.069**	2.627	2.376	1.860	1.583	1.450
		1,500	5.193	2.788	1.756	1.398	**1.198**	1.810	1.256	1.082	**1.082**	**1.057**	2.869	2.219	1.623	1.292	**1.133**

Table 18 (continued)

List Size	πs	n	Percent of Truthful Reporting														
			Ratio of MSE: ICT/DICT					Ratio of MSE: SSC/DICT					Ratio of MSE: ICT/SSC				
			1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6	1.0	0.9	0.8	0.7	0.6
	0.25	150	5.366	4.780	3.133	2.572	2.035	1.906	1.779	1.471	**1.294**	**1.334**	2.815	2.687	2.129	1.988	1.526
		500	5.678	3.408	2.043	1.610	1.483	1.995	1.469	**1.270**	**1.116**	**1.120**	2.846	2.320	1.609	1.442	1.324
		1,500	5.235	2.351	1.464	**1.184**	**1.115**	1.911	1.371	**1.041**	**1.027**	**1.026**	2.739	1.714	1.406	1.153	**1.087**
	0.35	150	5.384	3.735	2.627	1.928	1.557	1.870	1.578	1.357	**1.190**	**1.158**	2.880	2.367	1.936	1.619	1.345
		500	5.529	3.032	1.705	**1.323**	**1.241**	1.945	1.449	**1.153**	**1.088**	**1.047**	2.843	2.092	1.479	**1.216**	**1.185**
		1,500	5.605	1.899	1.242	**1.154**	**1.077**	2.068	**1.166**	**1.055**	**1.031**	**1.009**	2.710	1.628	1.176	**1.119**	**1.068**
	0.45	150	4.394	3.481	2.211	1.568	1.424	1.623	1.547	**1.265**	**1.024**	**1.057**	2.707	2.250	1.747	1.532	1.348
		500	4.841	2.492	1.565	**1.186**	**1.156**	1.760	**1.272**	**1.169**	**1.058**	**1.046**	2.750	1.959	1.339	**1.121**	**1.105**
		1,500	5.114	1.671	1.141	**1.053**	**1.030**	1.907	**1.193**	**1.012**	**1.012**	**1.026**	2.681	1.401	**1.127**	**1.040**	**1.003**

Table 18 indicated that the efficiency patterns were similar across all list sizes (3-item, 4-item, 5-item). In general, the double item count technique (DICT) optimal model was more efficient (i.e., ratio of MSE $\geq$ 1) than the single sample count technique (SSC) optimal and item count technique (ICT) optimal models across all sample sizes, sensitive prevalent rates and non-compliance rates. In addition when comparing the ICT optimal model to the SSC optimal, the SSC was generally more efficient (i.e., ratio of MSE $\geq$ 1) than the ICT optimal across all sample sizes, sensitive prevalent rates and non-compliance rates. Thus, results from this study indicated that generally the optimal DICT model proved to be more efficient than the optimal SSC and ICT models, especially when estimating smaller to mid-size sensitive prevalent rates (i.e., $0.01 \leq \pi_s \leq 0.15$) across all sample sizes.

CHAPTER V

DISCUSSION

Selecting the Optimal Model

The idea of finding optimal models for these analyses was to minimize variability so that the models selected for the compliance and list-size analyses would be the most efficient. As a result, using the relative reliability, the variation of each technique for all simulated parameters were compared to the variation of the direct questioning technique (DQT), under the assumption of truthful reporting, since it was the most efficient. By doing this, the parameters of each technique selected ensured the variability of the model was as close to that of the DQT as possible. For this portion of the analysis, direct comparisons of each technique to the DQT were not made nor discussed. Recall that the random response (RR) and non-random response (NRR) techniques were developed as an attempt to improve honest reporting of sensitive behaviors since it was suspected that estimates of these behaviors using the DQT were extremely biased. As a result, there was no need to examine how each technique compared to the DQT especially under the assumption of truthful reporting since, if it were expected that respondents report honestly, the most efficient technique to use would be the DQT.

The results of this analysis indicated that the unrelated question technique (UQT) model with parameters $p_1 = 0.90$, $p_2 = 0.10$, and $\pi_{ns} = 0.10$ proved most efficient against all other simulated UQT model parameters. This is not surprising since Greenberg et al. (1969) indicated that the variability of the UQT was minimized when the probability of selecting the sensitive question for both samples were as far from 0.50 as possible and at the same time, as far apart from one another as possible. Additionally, the authors also pointed out that the selection of a non-sensitive behavior with low prevalence also reduced the variability of the unrelated question technique (UQT). The set of simulations performed in this study confirmed these assumptions. Moreover, the results demonstrated that the choice of p_1 and p_2 were especially important since these parameters were more influential on the variation of the UQT compared to the prevalence rate of the non-sensitive behavior (i.e., π_{ns}). Figure 1 illustrated this by plotting the relative reliability of the UQT model as it compared to the variability of the DQT for sensitive prevalent rates meant to represent small ($\pi_{ns} = 0.05$) medium ($\pi_{ns} = 0.20$), and large ($\pi_{ns} = 0.45$). As is evident in the plot, where the figure peaks, the UQT model is closest to the variability of the DQT model and is more efficient. When the plot valleys, this indicated that the selected model parameters of the UQT increased model variability and the UQT was less efficient compared to the DQT. As was evident in the plot, as p_1 decreases from 0.90 to 0.60, the plot of the relative reliability decreased substantially, indicating that model variation had increased. Also evident in the plot, the variability of the UQT also increased as the prevalence of the non-sensitive behavior increased, but the increase to the variance was not as substantial. This is important to know, since researchers who utilize the UQT model need to be aware that their choice of how frequently the sensitive question was asked (i.e., p_1, p_2) in each sample impacts the models variation compared to the selection of the non-sensitive behavior (i.e., π_{ns}).

For the forced choice technique (FCT), model parameters were already selected and based on the sum of two dice, where probabilities of selecting the sensitive question, forcing the respondent to report “no” or “yes” were pre-determined. This model was selected since previous studies that investigated non-compliance used this version of the FCT (Böckenholt et al., 2009; Böckenholt & van der Heijden, 2004; van den Hout et al., 2010). As a result, the FCT model was not studied for optimality in this analysis.

For the NRR models, the results of this analysis were especially important since they test the assumptions made by Glynn (2013) in determining an optimal NRR. In order to simultaneously minimize ceiling effects and response variability without compromising privacy, Glynn (2013) suggested a method defining an optimal design for the NRR. First, equally allocate subjects into the two groups since as demonstrated by Glynn (2013) equal sample sizes can actually benefit the design--especially in terms of the double list technique. For the set of simulations, sample sizes were equally distributed between the two samples for the item count technique (ICT) and double item count technique (DICT). Since the single sample count (SSC) only used one sample of participants this was not necessary. According to Glynn (2013), a more potential method in reducing variation was the selection of the innocuous questions, their prevalence rates and how the items correlated. Recall that Glynn (2013) suggested avoiding a high number of non-sensitive questions with either high or low prevalent rates and in addition to select pairs of innocuous questions that correlate negatively. This would reduce the likelihood of ceiling and floor effects as well as variability since the number of “how many” items a respondent reports decreases. In addition, for the DICT model, since two lists were administered, if they were highly correlated, variability can further be reduced (Glynn, 2013). These assumptions were tested in the set of simulations that determined the optimal model for each NRR. Again, truthful reporting was assumed and NRR models were compared to the DQT for each of the sensitive prevalent rates, sample size and item list size combinations, in essence to select the model closest to the

DQT in efficiency. For the item count technique (ICT) and double item count technique (DICT) the distribution of the prevalent rates of the non-sensitive questions were explored by testing three distribution types: “equal,” “symmetric but not equal,” and “not symmetric and not equal.” These were meant to explore Glynn’s suggestion of avoiding in abundance non-sensitive questions with either high or low prevalent rates and followed previous studies that explored the effects of differing distribution rates of non-sensitive questions in an item list (Blair & Imai, 2012; Corstange, 2009). For the single sample count technique (SSC), this again was not necessary since the prevalent rates of the non-sensitive questions were selected as 0.50 by design. In addition, simulations also explored the affects of negatively correlated pairs of non-sensitive questions in each item list as it pertains to statistical efficiency. This was done for the NRR models, the ICT and SSC. For the DICT, simulations that adjusted for both *within* and *between* list correlations could not be made. As a result, a simulation study using the DICT with the highest variability (i.e., the 5-item DICT estimating $\pi_s = 0.45$) determined that the *between* list correlation was substantially more effective in reducing model variability compared to the *within* list correlation. Thus, for this study, two sets of simulations were performed for the ICT, SSC, and DICT in an effort to demonstrate the claims made by Glynn (2013). The first adjusted for *within* list (ICT, SSC) or *between* list correlation (DICT) arbitrarily selected at -0.50 and 0.85, respectively. The second set of simulations did not adjust for *within* (ICT, SSC) correlation between non-sensitive questions in the item list. Optimal models were then selected, by sensitive prevalent rate and sample size combination, if the efficiency of the model was closest to the efficiency of the DQT.

The results of this analysis revealed that for each NRR model, correlating between sensitive questions *within* a list (ICT, SSC) or *between* lists (DICT) effectively reduced model variation as postulated by Glynn (2013). For each NRR model, a model that correlated *between* or *within* a list was selected as optimal. In addition, in examining the distribution of prevalent rates of the non-sensitive questions (i.e., “equal,” “not equal but symmetric,” and “not equal and not symmetric”) the results of this analysis indicated that distribution type is more effective in reducing variation for the item count technique (ICT) compared to the DICT. In the former case, reduction in model variability by the distribution of non-sensitive questions were consistent across all sensitive prevalent rates and sample size combination and differed by list size. For the ICT 3-item and the ICT 5-item models, distributions of non-sensitive questions that were not equal reduced variation compared to the corresponding models with equally distributed prevalent rates of non-sensitive questions. For the ICT 3-item, the “not equal and Not symmetric” distribution type was selected as the optimal model whereas for the ICT 5-item, “the symmetric and not equal” was selected as the optimal model. As a result, in general a distribution of non-sensitive items that are “not equal” appeared to reduce additional variation in the ICT 3-item and 5-item models. For the ICT 4-item, when the prevalence rate of non-sensitive questions are distributed equally, the model was more efficient in reducing variation compared to the ICT 4-item models where distribution of non-sensitive prevalent rates were “not equal.” It is important to keep in mind that within each of these list items, pairs of sensitive questions were also negatively correlated at -0.50 and this correlation also influenced the model’s variation. For the 3-item list, only one pair of questions could be negatively correlated whereas for the 5-item lists 2 pairs of questions could be correlated. For the 4-item list, correlation between pairs of non-sensitive questions were made based on the distribution type (i.e., “not equal but symmetric,” “not equal and not symmetric”) of the prevalent rates of the non-sensitive questions. This again was due to the constraints by the marginal probabilities on the allowable correlation (Leisch et al., 1998). Thus, for the 4-item list when the distribution of non-sensitive questions was unequal, only one pair of non-sensitive questions could be correlated. The reduction of variability in the item count technique (ICT), therefore, appeared to be influenced by the combination of both the distribution of prevalent rates of the non-sensitive questions and the number of negatively correlated pairings of non-sensitive questions in the item list. For the 4-item list size, since two pairs of non-sensitive questions could be negatively correlated if the distribution of the prevalent rate of non-sensitive questions was equal, compared to just one pair of negatively correlated non-sensitive questions when the prevalent rate distribution was not equal, the correlation appeared to have influenced the reduction in variation since the former model (“equal”) was selected as optimal. For list sizes that were odd (i.e., 3-item and 5-item), since all but one of the non-sensitive questions could be correlated, the distribution of the prevalent rates of the non-sensitive questions played a larger role in further reducing model variability. For both list sizes, correlations between non-sensitive items could be made for the maximum number of pairings (i.e., 1 pair in the 3-item model, 2 pairs in the 5-item model) across all non-sensitive prevalent rate distribution types (“equal,” “not equal but symmetric,” and “not equal and not symmetric”). In examinimg the selected optimal models for both list sizes, the non-sensitive question left uncorrelated in both the 3-item and 5-item model as defined in the simulation study, had the lowest prevalent rate (i.e., 1/4th for both the “not equal and symmetric” and “not equal and not symmetric” in the 3-item model; 1/6th and 1/7th for the “not equal and symmetric” and “not equal and not symmetric” in the 5-item model). Thus, because the prevalent rate of the uncorrelated question was small, the number of “yes” responses to the non-sensitive questions were reduced which decreased the size of the sum across all questions in the list subsequently reducing model variation.

Therefore, according to the results of this study, for the ICT the distribution of prevalent rates of non-sensitive questions in the item list as well as pairing of correlated non-sensitive questions both played a role in reducing model variation. For the double item count technique (DICT), the results of the study determined that the correlation *between* item lists more effectively reduced model variation compared to the correlation *within* list items or the distribution of the non-sensitive prevalent rates in the item list. In the former case, the relative reliability analysis that compared efficiency between DICT models with the highest expected variation (i.e., DICT 5-item, $\pi_s = 0.45$, $n = 150$) when *between* and *within* item lists were left uncorrelated, selected *within* item lists were correlated at -0.50 and *between* list items were correlated at 0.85, the *between* list correlations were substantially more efficient. Therefore, the results of this study indicated that when *between* list correlations were high in the DICT, model variation was substantially reduced. Likewise for the SSC, the results of this study indicated that if pairs of non-sensitive questions were negatively correlated, model variation was also reduced.

In determining the optimal models for the both RR and NRR models, this study confirmed the assumptions made by both Greenberg et al. (1969) and Glynn (2013). For the UQT model, by optimizing model parameters, variability was reduced and models were more similar to the DQT under the assumption of truthful reporting. Likewise, for the NRR models, negatively correlating non-sensitive questions *within* item lists for the item count technique (ICT) and single sample count technique (SSC) resulted in more efficient models; whereas positively correlating all items *between* lists in the double item count technique (DICT) reduced variation and in turn increased efficiency. These results were consistent across all sensitive prevalent rates and sample size combinations. Thus, in accordance to the results of this study, researchers who use RR and NRR models to estimate sensitive prevalent rates are now able to optimize the design of their survey. For those who choose to use the UQT, the probability of selecting the sensitive question in each sample (i.e., p_1, p_2) is especially important since this parameter affects the variability of the model more substantially than the choice of the non-sensitive prevalent rate (i.e., π_{ns}). For the NRR models, the results of this study indicated that correlating *within* item lists (ICT, SSC) or *between* item lists (DICT) reduced the variation of the model. Therefore, researchers should think carefully about the non-sensitive questions that make up the item list and consider negatively correlating pairs. In addition, if using the ICT, the distribution of prevalent rates of the non-sensitive questions should also be considered. In odd numbered lists, if only a selection of pairs of non-sensitive items were negatively correlated, sensitive prevalent rates distributed as not equal reduced model variation where the item left uncorrelated should have a lower prevalent rate. For the even item list if all non-sensitive items were negatively correlated in pairs, the distribution of the prevalent rates of the non-sensitive items reduced model variability when they were equal.

Study of Non-compliance: Random Response vs. Non-random Response and the Effects of List Size

The results of this study indicated that the random response model, the unrelated question technique (UQT) optimal was generally more efficient (i.e., ratio of MSE $\geq$ 1) than the NRR models, optimal DICT, SSC and ICT across all sample sizes, sensitive prevalent rates and equivalent non-compliance rates. In addition, the double item count technique (DICT) optimal 3-item and 4-item models were as efficient as the forced choice technique (FCT) model across all sample sizes, sensitive prevalent rates and non-compliance rates. For the UQT practical model, the NRR models were generally more efficient (i.e., ratio of MSE $\geq$ 1). In the comparison of each non-random response (NRR) model to each random response (RR) model, a similar pattern of the ratio of MSE emerged where the DICT optimal 3-item, 4-item, and 5-item models proved most efficient compared to the optimal single sample count technique (SSC) and item count technique (ICT) models since, when comparing these models to the corresponding RR models, the ratios of MSE were generally larger for the DICT models. The SSC optimal models, however, were more efficient than the ICT optimal models, which proved to be the least efficient of all the NRR techniques.

The effects of list size on statistical efficiency among the optimal NRR models indicated that in general, across all sensitive prevalent rates, sample sizes and equivalent non-compliance rates, the 3-item and 4-item list sizes were similar in terms of efficiency whereas the 3-item list size proved more efficient than the 5-item list size. In terms of efficiency between NRR models, in general, the optimal DICT models proved most efficient (i.e., ratio of MSE $\geq$ 1) compared to both the optimal SSC and ICT models across all sample sizes, sensitive prevalent rates, non-compliance rates and list sizes. In turn, the optimal SSC was also more efficient compared to the optimal item count technique (ICT) models across all parameters and list sizes. However, for larger sensitive prevalent rates and increasing sample sizes, statistical efficiency between 3-item and 5-item NRR models approached unity for increasing levels of non-compliance.

In general, the results indicated that for all comparisons made in the study of non-compliance, the measure of statistical efficiency through the ratio of MSE approached unity as sensitive prevalent rates, non-compliance rates and sample sizes increased. This approaching efficiency was primarily due to the two components of the MSE, the variance and the bias. The MSE is generally a measure of the goodness or closeness of an estimator to its estimate. (Mood, Graybill, & Boes, 1950). Since the MSE is the sum of the estimator's variance and the square of its bias, when bias increases, its contribution to the MSE mounts rapidly, indicating that the estimator was further from the actual estimate. When comparing two estimators using the ratio of MSE, therefore, if the ratio of MSE was close to unity, one concludes that the models were similar in terms of each estimator's closeness to the actual estimate. In this study, bias was defined by the rate of compliance which ranged between 90% truthful reporting to 60% truthful reporting. Because this study examined the effects of non-compliance for a range of sensitive prevalent rates (π_S: 0 .01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.36, and 0.45) and sample sizes (n = 150, n = 500, and n = 1,500), these factors affected the magnitude of the bias and in turn the results of the study. When sensitive prevalent rates were small (π_s = 0.01 - π_s = 0.05), the resulting bias was generally less than the bias of larger sensitive prevalent rates. For example, if 60% of the respondents reported truthfully when the sensitive prevalent rate is 0.05, the resulting square of the bias was 0.0004. However, if 60% of the respondents reported truthfully when the sensitive prevalent rate was 0.45, the resulting square of the bias was 0.0324, a much larger value. Therefore, in the presence of high non-compliance, the contribution of the bias to the MSE was particularly influential when estimating larger sensitive prevalent rates compared to smaller sensitive prevalent rates. At the same time, sample size also influenced the MSE since larger sample sizes generally reduce model variation. As a result, when estimates are biased, as sample sizes increase and variation decreases, the contribution of the bias to the MSE also increases and becomes more influential. Thus, for larger sensitive prevalent rates with larger sample sizes, in the presence of high non-compliance, the MSE was primarily a measure of the squared bias. To provide an example of this, for the unrelated question technique (UQT) optimal when $\pi_s = 0.05$ and $n = 150$, if 60% of participants with the sensitive attribute reported honestly, the contribution of the square of the bias to the MSE was 33%. When sample sizes are increased to $n = 1{,}500$, the contribution to the square of the bias of the MSE for UQT optimal increased to 83%. When $\pi_s = 0.45$ and $n = 150$, if 60% of participants with the sensitive attribute report honestly, the contribution of the square of the bias to the MSE for the UQT optimal was 95% and increased to 99% when sample sizes were increased to $n = 1{,}500$. For the double item count technique (DICT) optimal 3-item, when $\pi_s = 0.05$, and $n = 150$, if 60% of participants with the sensitive attribute reported honestly, the contribution to the square of the bias of the MSE was 19%. When sample sizes were increased to $n = 1{,}500$, the contribution of the square of the bias to the MSE for DICT optimal 3-item increased to 73%. When $\pi_s = 0.45$ and $n = 150$, if 60% of participants with the sensitive attribute reported honestly, the contribution of the square of the bias to the MSE for the DICT optimal 3-item was 93% and increased to 99% when sample sizes were increased to $n = 1{,}500$. As a result, since in this study bias was equivalent in both models, when comparing between the models for $\pi_s = 0.45$ and $n = 150$ or $n = 1{,}500$, since the ratio of MSE approached unity the models were termed efficiently similar due to the fact that the estimates from both models were so biased.

As was seen in Figure 1, the variance of the UQT was influenced heavily by the selection of p_1 and p_2. In the optimal model, these parameters were selected as far from one another as possible (Sample 1: $p_1 = 0.90$, $p_2 = 0.10$, Sample 2: $p_1 = 0.10$, $p_2 = 0.90$) which in turn minimized the variance of the unrelated question technique (UQT). Therefore, the variance of the UQT optimal model was especially efficient. For the UQT practical, the selection of p_1 and p_2 were not optimal and in fact were made as close to 0.50 as possible, which in turn maximized the variance of the UQT. Thus, the variance of the UQT practical model was less efficient. When comparisons were made between the UQT models with each NRR optimal, for smaller sensitive prevalent rates, since the bias was less influential to the MSE, the techniques with smaller model variability proved more efficient (i.e., optimal UQT vs. all optimal NRR models, all optimal NRR models vs. UQT practical). As sample sizes and sensitive prevalent rates increased, for higher levels of non-compliance, the bias of the estimate heavily influenced the MSE and models were termed more similar--due to the bias. For the forced choice technique (FCT) a similar pattern was noted, where the FCT and DICT optimal 3-item model were more efficient when sensitive prevalent rates were smaller but became more similar for increasing sensitive prevalent rates, sample sizes and non-compliance. Likewise, when comparing NRR models by list size to one another, similar patterns were also noted.

Thus, for NRR optimal and RR models used to estimate high sensitive prevalent rates with larger sample sizes, if high rates of non-compliance were present, models were primarily similar in terms of their bias.

Because bias largely influenced the MSE in cases of larger sensitive prevalent rates and sample sizes when non-compliance rates were equivalent, it is difficult to assess the effects of non-compliance when comparing models at these parameter levels. As a result, secondary analyses were made that explored the statistical efficiency between selected models when compliance rates differed. For the study exploring effects of non-compliance between RR models and NRR models, the unrelated question technique (UQT) optimal and double item count technique (DICT) optimal (all list sizes) were selected since, from this study, both models were the most efficient of their model class (i.e., RR vs. NRR). For the study exploring statistical efficiency between list sizes for each NRR model, the 3-item and 5-item models were selected to determine if there were situations in which, if the 5-item model improved compliance rates, it would be more efficient. For the analysis comparing the UQT optimal to the DICT optimal, the results indicated that for moderate to large sensitive prevalent rates ($\pi_s = 0.20$, $\pi_s = 0.45$), across list size (3-item, 4-item, and 5-item) and sample size ($n = 150$, $n = 500$, and $n = 1{,}500$), if the DICT optimal improved the compliance rate by 10% or more, the DICT optimal was just as or more efficient (i.e., ratio of MSE ≥ 1). Efficiency increased between these models in the presence of differing non-compliance as sample sizes increased. For smaller sensitive prevalent rates ($\pi_s = 0.05$), when sample sizes are large ($n = 1{,}500$), if the DICT optimal improved the compliance rate by 20% or more, the DICT optimal was more efficient. For the NRR comparisons between 3-item and 5-item models, results indicated that across all NRR models and sample sizes when sensitive prevalent rates were moderate or large ($\pi_s = 0.20$, $\pi_s = 0.45$), if the 5-item model improved compliance by at least 10% from the 3-item model, the 5-item model was just as or more efficient (i.e., ratio of MSE $\geq$ 1.00). If improvement in compliance was 20% or more, the 5-item optimal double item count technique (DICT) and single sample count technique (SSC) were nearly twice as efficient as the corresponding 3-item model for moderate to larger sensitive prevalent rates. In addition, for smaller sensitive prevalent rates ($\pi_s = 0.05$), if the DICT 5-item optimal improved compliance rates by 20%, when sample sizes were large (n = 1,500), the 5-item model was more efficient (i.e., ratio of MSE $\geq$ 1.00)

These results suggested that for moderate sensitive prevalent rates, across all samples sizes, if the DICT 5-item optimal model improved compliance, estimates were more efficient than the unrelated question technique (UQT) optimal. For smaller prevalent rates, when sample sizes were large (n = 1,500), the DICT 5-item list was more efficient if it improved substantially the amount of truthful reporting.

These assumptions are not far fetched since the design of the UQT optimal was more prone to reduce model variability than increase confidentiality; whereas the design of the DICT 5-item model was meant to increase confidentiality at the cost of additional model variability. Recall that for the UQT optimal model, subjects randomized to the first sample responded to the sensitive question 90% of the time. In addition, since the prevalent rate of the non-sensitive question is low (i.e., $\pi_{ns} = 0.10$), if the sensitive attribute was more prevalent (i.e., $\pi_s > 0.10$), participants with the sensitive trait could become suspicious since they would be more likely to respond “yes” to the sensitive question compared to the non-sensitive question. As pointed out by Greenberg et al.

(1969), although the UQT optimal minimized model variation at the same time the design parameters could potentially reduce confidentiality resulting in higher bias. On the other hand, NRR models were designed to increase confidentiality by embedding the sensitive question in a list of non-sensitive questions. Since the respondents only report the number of items from the list they endorse, participants never directly respond to the sensitive question. Confidentiality was further protected if the item list is longer since the probability of responding "yes" to all questions in a shorter item list was less likely to occur than responding "yes" to all questions in a longer item list. Therefore, it is not difficult to assume that the non-random response (NRR) double item count technique (DICT) 5-item optimal model could improve compliance compared to the optimal unrelated question technique (UQT) or 3-item DICT optimal.

In estimating smaller sensitive prevalent rates, the results of this study were also useful since it suggested that if the double item count technique (DICT) optimal model can substantially improve compliance, estimates from these models were more efficient only when sample sizes were very large (n = 1,500). Since sensitive prevalent rate that are smaller, generally have less members, encouraging honest responding is essential in these populations. Also due to the small prevalence of the group, it is well known in sampling theory that drawing samples from these populations at higher rates improved efficiency. Thus, models that provide higher levels of confidentiality by encouraging honest reporting would be more useful for smaller populations if sample sizes were adequate. Since the UQT optimal was a "lower confidentiality" model, and the DICT 5-item optimal was a "higher confidentiality" model, the results of this analysis suggested, that for large sample sizes, if the DICT 5-item improved compliance substantially (i.e., by 20% or more), than the DICT 5-item model resulted in more efficient estimates compared to the UQT optimal. Results of this study, however, also suggested that for smaller sample sizes, the variability of the NRR models in estimating smaller sensitive prevalent rates proved inefficient compared to the UQT optimal.

In addition, the results of this study determined that non-random response (NRR) DICT optimal models were generally more efficient than the SCC and item count technique (ICT) optimal models and that the single sample count technique (SSC) optimal model was generally more efficient than the ICT optimal model. In fact, as shown by the results of this study, the ICT optimal models were the least efficient of all models with the exception of the unrelated question technique (UQT) practical. This was because both the double item count technique (DICT) and single sample count technique (SSC) optimal models estimated the sensitive prevalent rate using the entire sample of subjects since all subjects in these models responded to a list of questions containing the sensitive question. For the ICT, on the other hand, only one sample responded to a list containing the sensitive question. As a result, since a higher number of participants responded to the sensitive question when the DICT or SSC is utilized, the estimate of the sensitive prevalence rate was more efficient compared to the ICT that only utilized half the number of participants in estimating the sensitive prevalent rate. This inefficiency was apparent though out the study, where generally the ICT optimal models were less efficient than the UQT optimal, FCT, DICT optimal, and SSC optimal. The model that proved to be the most inefficient, however, was the UQT practical model since the parameters selected for this model maximized the variation.

Summary and Guidelines

This study was the first to explore statistical efficiency between random response models (RR) and non-random response models (NRR) in an extensive simulation study that also examined the effects of non-compliance on estimation using these techniques, the effects of list size for NRR models and efficiency between NRR models. The study was extensive, encompassing a larger range of sensitive prevalent rates than previous studies. The effects of sample sizes were also included where samples were arbitrarily categorized as small (n = 150), medium (n = 500), and large (n = 1,500) and studied in the set of simulations. In addition, the study also focused on determining optimal model parameters for the unrelated question technique (UQT) and NRR models that would successfully reduce variation prior to studying the effects of non-compliance. The results of this study were especially important since they verified the assumptions made by Greenberg et al. (1969) and Glynn (2013). This was especially apparent for the NRR models, in which case has never been formerly explored, where the effects of correlating *within* (ICT, SSC) and *between* (DICT) non-sensitive items in the item list were especially noteworthy in reducing model variability. This is also the first extensive simulation study to explore the effects of non-compliance on estimates between RR and NRR models. The study revealed situations where if the non-random response (NRR) DICT optimal models can improve compliance rates compared to the UQT optimal, the DICT optimal was more efficient. Effects of NRR list sizes as it pertains to statistical efficiency in the presence of truthful and non-truthful reporting were also explored. In general, when the 5-item list improved compliance rates compared to 3-item lists, the 5-item list proved more efficient for the optimal DICT and SSC models when sensitive prevalent rates were moderate ($\pi_s = 0.20$) or large ($\pi_s = 0.45$), across all sample sizes. The study also examined statistical efficiency between NRR models and indicated that the techniques utilizing the entire sample to estimate the sensitive prevalent rate (i.e., DICT optimal and SSC optimal) proved more efficient compared to the model that utilized only half the sample of subjects to estimate the sensitive prevalent rates (i.e., ICT). Lastly, this was also the first study to provide an official definition of non-compliance for NRR models in terms of under-reporting the sensitive trait.

In summary, based on the results of this study, the following guidelines and recommendations were developed:

1. In order to effectively implement the unrelated question technique (UQT) model, careful consideration of the rate at which the sensitive question is asked in both samples should be thought through since this rate was more influential on model variability compared to the selection of the non-sensitive behavior (i.e., π_{ns}). Sample sizes and probabilities of selecting the sensitive question in both samples (i.e., p_1, p_2) should be selected as outlined by Greenberg et al. (1969), where p_1 (the probability of selecting the sensitive question for sample 1) and p_2 (the probability of selecting the sensitive question for sample 2) should be as far from 0.50 as possible and have the same but complementary effect (i.e., $p_1 + p_2 = 1$) in each sample. In terms of sample sizes, subjects should be optimally allocated into samples based on estimates of the sensitive prevalent rate as described by Greenberg et al. (1969).

2. In order to effectively reduce the variation present in NRR models (i.e., ICT, DICT, and SSC), the suggestions of Glynn (2013) should be followed. For the DICT, highly correlated lists reduced model variation substantially; whereas negatively

correlating pairs of non-sensitive items within a list reduced model variation of the item count technique (ICT) and single sample count technique (SSC). In addition, depending on the number of paired correlations in an item list, for the 3-item and 5-item ICT, if the distribution of the non-sensitive items were made not equal, this would further reduce model variation. For the 4-item list, evenly distributed prevalent rates of non-sensitive behaviors in the item list reduced variation when all pairs of questions were correlated.

3. The unrelated question technique (UQT) optimal model was the most efficient of the random response (RR) and non-random response (NRR) models compared in this study (i.e., FCT, ICT, DICT, and SSC) if the expected amount of truthful reporting is equivalent between models. If it is expected that any of the double item count technique (DICT) optimal models (i.e., item list sizes 3, 4, or 5) improved compliance rates by more than 10%, for moderate (i.e., π_s = 0.10, 0.15, and 0.20) to larger (i.e., π_s = 0.25, 0.35, and 0.45) sensitive prevalent rates across all sample sizes (*n* = 150, *n* = 500, and *n* = 1,500), the DICT 5-item optimal model was more efficient. For smaller sensitive prevalent rates (i.e., π_s = 0.01, 0.03, and 0.05), if it is expected that the DICT 5-item optimal model would improve compliance rates substantially (i.e., greater than 20%), when sample sizes are large (*n* = 1,500), the DICT 5-item optimal model was more efficient.

4. In general, for the DICT optimal and SSC optimal, 3-item list sizes were more efficient in estimating sensitive prevalent rates unless the 5-item list size improved compliance rates. In the case of moderate (i.e., π_s = 0.10, 0.15, and 0.20) to larger (i.e., π_s = 0.25, 0.35, and 0.45) sensitive prevalent rates, these 5-item optimal models were more efficient than the 3-item model. For smaller prevalent rates (i.e., π_s = 0.03, 0.05), for the

DICT optimal model exclusively, if the 5-item model improved compliance rates substantially (i.e., $\geq$ 20%) then the 5-item model was more efficient when sample sizes were large (n = 1,500).

5. In general, the item count technique (ICT) optimal and unrelated question technique (UQT) practical models were the least efficient NRR models. The DICT optimal model where non-sensitive item lists were highly correlated proved the most efficient of the NRR models.

Limitations and Future Research

There are several limitations of this study. For one, since the study of Greenberg et al. (1969) was followed, sample sizes for the unrelated question technique (UQT) were optimally allocated as the presence of non-compliance increased. In real world application, this would suggest that the researcher is aware of the non-compliance rate prior to the study and therefore is able to effectively allocate the sample. As a result, since the UQT optimal proved to be the most efficient in this study, an additional analysis where sample sizes are allocated based on the estimated sensitive prevalent rate and then held fixed at these levels should be studied to confirm these results. A second limitation of the study occurred with the single sample count technique (SSC) where it was assumed that the prevalent rates of the non-sensitive questions are fixed at 0.50. In reality, unless the prevalent rates of the non-sensitive questions were exactly 0.50, additional variation due to the estimation of the non-sensitive questions should be accounted for in the simulation study--especially since additional variation would reduce statistical efficiency. A third limitation of the study is due to the arbitrary selection of the correlation rates for the *within* and *between* list correlations (i.e., -0.50, 0.85). These were selected for consistency since pairwise correlations could be simulated at these levels within each non-sensitive prevalent rate distribution type (i.e., “equal,” “not equal and symmetric,” and “not equal and not symmetric”) selected for this study. Based on these distributions, for the ICT 4-item list, when the distribution of the non-sensitive items was not equal, this restricted the number of allowable correlated pairs, since as discussed previously allowable correlations between two random variables were restricted by their marginal probabilities (i.e., in this case the marginal probabilities were the prevalent rates of the pair of non-sensitive questions). Because of this, changes to the probability distribution of the prevalent rates of non-sensitive items that allow for higher *within* list correlation could be studied in terms of increasing efficiency in the ICT. In addition, by restricting the *within* correlation to -0.50, the results of the single sample count technique (SSC) may have been undermined especially as it compares to the double item count technique (DICT). In a future study, an optimal correlation specific to the SSC could be explored and compared with an optimal DICT to determine if the SSC can be made as efficient. In addition, the arbitrary selection of the *between* list correlation for the DICT may have given this model an advantage since each of the non-sensitive questions could be highly correlated at 0.85. For the ICT and SSC, non-sensitive items were restricted to a smaller correlation (i.e., -0.50). Future studies, therefore should be attempted to determine optimal non-sensitive prevalent rate distributions and corresponding *within* and *between* list item correlations for the ICT and DICT in reducing model variation. In addition, simulations of the DICT that study effects on model variation due to both *within* and *between* list item correlations should also be explored. Once these optimal prevalent rate distributions and correlations are found, a similar simulation study of non-compliance could be performed and compared to the results of this study. In addition, the distribution of the prevalent rates for the non-sensitive item list in the ICT and DICT were also arbitrarily determined. These distributions were explored in previous simulation studies (Blair & Imai, 2012; Corstange, 2009). However, it might be of interest to empirically study particular lists from actual NRR surveys where prevalence rates of the non-sensitive items could be estimated and its effect on efficiency in the presence of non-compliance simulated and studied. In addition, by empirically studying item lists, as is done in the development of surveys, reliable manufactured lists could be developed for each of the NRR models where predefined correlations and non-sensitive prevalent rate distributions are known. Once these lists are developed, a similar simulation study would be performed using these model parameters to study more effectively the effects of non-compliance as it pertains to lists of real world applications. Fourthly, a further limitation of the study is the manner in which non-compliance was defined. The definition of non-compliance was restricted to subjects with the sensitive trait in which misreporting only occurred when these subjects were asked the sensitive question. Otherwise, subjects were assumed to report truthfully. This restricted definition of non-compliance followed the study of Greenberg et al. (1969). However, as is indicated by the literature, non-compliance also occurred with innocuous questions. For one, subjects who were forced to respond "yes" when the forced choice technique (FCT) was used, may be reluctant to comply with the rules especially if the subject was extremely sensitive, and instead respond "no." This in turn results in a lower than expected number of "yes" responses, which results in additional bias. For the UQT technique, if subjects have control of the survey as was seen in the Petróczi et al. (2011) study, even though they were instructed to answer the sensitive question, they may instead choose to respond to the non-sensitive question. When this occurs, “yes” responses were potentially inflated and the probabilities associated with the selection of each question were no longer valid. As a result, the sensitive prevalent rate was over estimated. Finally, for the NRR techniques, participants who possess a higher number of non-sensitive traits in the item list may under report their membership if they also possess the sensitive trait. As a result, under reporting may also occur for the non-sensitive questions in the item list, which would further bias the estimate. Because non-compliance occurs for innocuous questions in each technique, a more complex definition of non-compliance should be developed and the effects explored in a similar simulation study. Lastly, the study was also limited since simulations of the NRR techniques relied on an algorithm that used an approximate normal distribution as a threshold in simulating binary data.

Conclusion

In conclusion, this was the first extensive study to examine statistical efficiency between random response (RR) and non-random response (NRR) models in the presence of non-compliance. It is also the first study to provide a definition of non-compliance for the NRR techniques. In doing so, the study was able to develop general guidelines meant to help researchers determine, under certain situations, which techniques produce estimates that are more efficient. In general, the results of this study indicated that the unrelated question technique (UQT) optimal model was the most efficient of the techniques in the presence of equivalent non-compliance rates. However, if the DICT optimal 5-item model improved compliance, this model became more efficient depending on the sensitive prevalent rate estimated and the sample size. As a result, the study was able to demonstrate that in certain situations, the non-random response (NRR) DICT optimal model was as or more efficient than the random response (RR) UQT optimal model.

REFERENCES

Ahart, A. M., & Sackett, P. R., (2004). A new method of examining relationships between individual difference measures and sensitive behavior criteria: Evaluating the unmatched count technique. *Organizational Research Methods, 7*(1), 101-114.

Belli, R. F., Traugott, M. W., & Beckmann, M. N. (2001). What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics-Stockholm, 17*(4), 479-498.

Biemer, P. P., & Wright, D., (2004). Estimating cocaine use using the item count methodology, in "Preliminary results form the national Survey on drug use and Health." *In Annual meeting of the American Association for Public Opinion research*, Phoenix, Arizona.

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis, 20*(1), 47-77.

Böckenholt, U., Barlas, S., & van der Heijden, P. G. M. (2009). Do randomized-response designs eliminate response biases? An empirical study of non-compliance behavior. *Journal of Applied Econometrics*, *24*(3), 377-392.

Böckenholt, U., & van der Heijden P. G. M. (2004). Measuring noncompliance in insurance benefit regulations with randomized response methods for multiple items. In A. Biggeri, E. Dreassi, C. Lagazio, & M. Marchi (Eds.), *19th international workshop on statistical modelling* (pp. 106-110). Florence, Italy: Sage Publications.

Böckenholt, U., & van der Heijden P.G.M. (2007). Item randomized-response models for measuring noncompliance: risk-return perceptions, social influences and self-protective responses. *Psychometrika, 72*(2), 245-262.

Boruch, R. F. (1971). Assuring confidentiality of responses in social research: A note on strategies. *The American Sociologist*, 6(4), 308-311.

Chaudhuri, A., & Christofides, T. C. (2007). Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference, 137*(2), 589-593.

Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: detecting cheating in the randomized response model. *Psychological Methods, 3*(2), 160-168.

Cobb, M. D. (2001). Unobtrusively measuring racial attitudes: *The consequences of social desirability effects* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Corstange, D. (2009). Sensitive questions, truthful answers? Modeling the list experiment with LISTIT. *Political Analysis, 17*(1), 45-63.

Cruyff, M. J. L. F., van den Hout, A., van der Heijden, P. G. M., & Böckenholt, U. (2007). Loglinear randomized-response models taking self-protective response behavior into account. *Sociology Method Research, 36*(2), 266-282.

Dalton, D. R., Wimbush, J. C., & Daily, C. M. (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. *Personnel Psychology, 47*(4), 817-829.

Dalton, D. R., Daily, C. M., & Wimbush, J. C. (1997). Collecting "sensitive" data in business ethics research: A case for the unmatched count technique (UCT). *Journal of Business Ethics, 16*(10), 1049-1057.

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. *Measurement errors in surveys*, 185-210.

Fox, J. P. (2005). Randomized item response theory models. *Journal of Educational and Behavioral Statistics, 30*(2), 189-212.

Fox J. A., & Tracy P. E. (1986). *Randomized response: A method for sensitive surveys*. Newbury Park, CA: Sage.

Fu, H., Darroch, J. E., Henshaw, S. K., & Kolb, E. (1998). Measuring the extent of abortion underreporting in the 1995 National Survey of Family Growth. *Family Planning Perspectives*, 128-138.

Glynn, A. N. (2013). What can we learn with statistical truth serum? design and analysis of the list experiment. *Public Opinion Quarterly, 77*(S1), 159-172.

Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association, 64*(326), 520-539.

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports Tests using the item count technique. *Public Opinion Quarterly, 74*(1), 37-67.

Horvitz, D. G., Shah, B. V., & Simmons, W. R., (1967). The unrelated question randomized response model. *Social Statistics Section Proceedings of the American Statistical Association*, *65*, 72.

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, *106*(494), 407-16.

Jann, B., Jerke, J., & Krumpal, I. (2012). Asking sensitive questions using the crosswise model: An experimental survey measuring plagiarism. *Public Opinion Quarterly, 76*(1), 32-49.

Kendall, M., & Stuart, A., (1979). *The advanced theory of statistics ed. 4*. London: England: Charles Griffin.

Kuk, A. Y. (1990). Asking sensitive questions indirectly. *Biometrika, 77*(2), 436-438.

Kuklinski, J. H., & Cobb, M. D. (1998). When white southerners converse about race. *Perception and Prejudice*, 35-57.

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the "new south." *Journal of Politics*, 323-349.

Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., & Mellers, B. (1997). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, 402-419.

Leisch, F., Weingessel, A., & Hornik, K. (1998). On the generation of correlated artificial binary data. *University of Economics and Business Administration*, Adaptive information systems and modeling in economics and management science, Working paper series, Vienna University of Economics. Available at http://www.wu-wien.ac.at/am (last accessed date December 30, 2008).

Lensvelt-Mulders, G. J., Hox, J. J., & van der Heijden, P. G. (2005). How to improve the efficiency of randomised response designs. *Quality and Quantity, 39*(3), 253-265.

Lensvelt-Mulders, G. J., Hox, J. J., Van der Heijden, P. G., & Maas, C. J. (2005). Meta-analysis of randomized response research thirty-five years of validation. *Sociological Methods and Research, 33*(3), 319-348.

Miller, J. D., (1984). *A new survey technique for studying deviant behavior* (Unpublished doctoral dissertation). George Washington University, Washington, DC.

Miller, J. D., Cisin, I. H., & Harrel, A. V. (1986, May). *A new technique for surveying deviant behavior: Item-count estimates of marijuana, cocaine, and heroin*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1950). *Introduction to the theory of statistics*. New York, NY: Mc-Graw Hill Book Company.

Ostapczuk, M., Moshagen, M., Musch, J., & Zhao, Z. (2010). Reducing socially desirable responses in epidemiologic surveys: an extension of the randomized-response technique. *Epidemiology, 21*(3), 379.

Pal, S., & Singh, S. (2012). A new unrelated question randomized response model. *Statistics, 46*(1), 99-109.

Petróczi, A., Tamás N., Cross, P., Taft, H., Shah, S., Deshmukh, N., . . . Naughton, D. P. (2011). New non-randomised model to assess the prevalence of discriminating behaviour: A pilot study on mephedrone. *Substance Abuse Treatment Prevention and Policy, 6*(1), 20.

Rayburn, N. R., Earleywine, M., & Davison, G. C. (2003). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment and Trauma, 6*(2), 137-152.

Rudas, T., Clogg, C. C., & Lindsay, B. G. (1994). A new index of fit based on mixture methods for the analysis of contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 623-639.

Saha, A, (2010). A modified unrelated question randomized response device for complex surveys. *Stat Papers, 51*(2), 349-355.

Shimizu, I. M., & Bonham, G. S., (1978). Randomized response technique in a national survey. *Journal of the American Statistical Association, 73*(361), 35-39.

Sniderman, P. M., & Grob, D. B. (1996). Innovations in experimental design in attitude surveys. *Annual review of Sociology*, 377-399.

Soeken, K. L., & Macready, G. B. (1982). Respondents' perceived protection when using randomized response. *Psychological bulletin, 92*(2), 487.

Thompson, S. K. (2002). *Sampling*. New York, NY: John Wiley and Sons Inc.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin, 133*(5), 859.

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

Tsuchiya, T. (2005). Domain estimators for the item count technique. *Survey Methodology, 31*(1), 41-51.

Tsuchiya, T., Hirai, Y., & Ono, S., (2007). A study of the properties of the item count technique. *Public Opinion Quarterly, 71*(2), 253-272.

van den Hout, A., Böckenholt, U., & van der Heijden, P. G. M. (2010). Estimating the prevalence of sensitive behavior and cheating with dual design for direct questioning and randomized response. *Applied Statistics*, 59(Part 4), 723-736.

van den Hout, A., & Klugkist, I. (2009). Accounting for non-compliance in the analysis of randomized response data. *Australian New Zealand Journal of Statistics, 51*(3), 353-372.

Warner S. L. (1965). Randomized response: a survey technique for eliminating answer bias. *Journal of the American Statistical Association, 60*(309), 63-69.

Wiseman, F., Moriarty, M., & Schafer, M. (1975). Estimating public opinion with the randomized response model. *The Public Opinion Quarterly, 39*(4), 507-513.

Wimbush, J. C., & Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82*(5), 756-763.

APPENDIX A

ACRONYMS

Acronyms

RR	Random response models
UQT	Unrelated question technique (Random response model)
FCT	Forced choice technique (Random response model)
NRR	Non-random response models
ICT	Item count technique (Non-random response model)
DICT	Double item count technique (Non-random response model)
SSC	Single sample count technique (Non-random response model)

APPENDIX B

R-CODE

R-Code: Determining number of simulations

```
#Used to determine number of sims - based on ICT and SSC
#Run initial check on variance to determine which model has highest variance
#This will be based on 200 iterations of 500 simulations selected arbitrarily.

*ICT - equal

NRRICTE <- function() {
#based on sample smallest sample size (150), smallest pi_s (0.01), no correlation, largest
list size (5)
#will determine which list (equal, unequal-sym,unequal - unsym)

#cori will use cbind function of define the correlation matrix for the intervention group
where there are no correlations between non-sensitive questions.

#corc will use cbind function of define the correlation matrix for the control group where
there are no correlations between non-sensitive questions.


cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0),c(0,0,0,0,1))

#define the probability distribution of non-sensitive questions: equal. Intervention group
will have additional probability for sensitive question.

ei<-c(2/3,2/3,2/3,2/3,2/3,.01)
ec<-c(2/3,2/3,2/3,2/3,2/3)

#Create binary lists for intervention & control group using the rmvbin function.

int<-rmvbin(75, margprob=ei, bincorr=cori)
cont<-rmvbin(75, margprob=ec, bincorr=corc)

#Sum across rows to create difference – in means estimate.

Sint<-rowSums (int, na.rm = FALSE, dims = 1)
Scontr<-rowSums (cont, na.rm = FALSE, dims = 1)

#Difference in means estimate

Pi_s <- mean(Sint) - mean(Scontr)
# output difference in means estimate
return(Pi_s)
}
```

```
##This function will run 200 iterations of 500 simulations selected arbitrarily calling the
function NRRICTE.

NRRICTEV <- function(n1, n2) {

#Create variance matrix for n1 variances.
Varmat <- matrix(nrow = n1, ncol=1)
        for (j in 1:n1) {
#create estimator matrix with n2 estimates calling NRRICTE function.
                        MAT<- matrix(nrow = n2, ncol=1)
                        for (i in 1:n2) MAT[i,] <-NRRICTE()
#Input n2 variances into var matrix.
#Input  variance of each iteration.
        Varmat[j,] <-var(MAT)
        }
#Take final variance of all n1 iterations.
        VarSim<-var(Varmat)
        return(VarSim)}

#IRT - EQUAL BUT SYMMETRIC
#This function is similar to the above function – only it uses the Equal but Symmetric
distribution.

NRRICTES <- function() {

#based on sample smallest sample size (150), smallest pi_s (0.01), no correlation, largest
list size (5)
#will determine which list (equal, unequal-sym,unequal - unsym)

#no corr

cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0),c(0,0,0,0,1))

#equal but symmetric distribution.

esi<-c(1/6,2/6,3/6,4/6,5/6,.01)
esc<-c(1/6,2/6,3/6,4/6,5/6)

#Create lists for intervention and control ICT groups.
int<-rmvbin(75, margprob=esi, bincorr=cori)
cont<-rmvbin(75, margprob=esc, bincorr=corc)

#Sum across rows to produce estimates.
Sint<-rowSums (int, na.rm = FALSE, dims = 1)
```

```
Scontr<-rowSums (cont, na.rm = FALSE, dims = 1)
#Create diff in means estimator.
Pi_s <- mean(Sint) - mean(Scontr)
return(Pi_s)

}

#This function will produce n1 variances based on n2 simulated diff of means estimates
using the unequal but symmetric distribution.

NRRICTESV <- function(n1, n2) {
#create variance matrix for n1 variances.
Varmat <- matrix(nrow = n1, ncol=1)
        for (j in 1:n1) {
#create estimator variance for n1 estimates.
                MAT<- matrix(nrow = n2, ncol=1)
                        for (i in 1:n2) MAT[i,] <-NRRICTES()
        #obtain variance from n1 simulations
        Varmat[j,] <-var(MAT)
        }
        #find variance from n2 variances.
        VarSim<-var(Varmat)
        return(VarSim)
}

#This function will be similar to previous functions – only will define for unequal and not
#symmetric.

NRRICTU <- function() {

#based on sample smallest sample size (150), smallest pi_s (0.01), no correlation, largest
list size (5)
#will determine which list (equal, unequal-sym,unequal - unsym)

#no corr

cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0),c(0,0,0,0,1))

#unequal and not symmetric – defined by user.
eusi<-c(1/7,3/7,4/7,5/7,5/7,.01)
eusc<-c(1/7,3/7,4/7,5/7,5/7)
#create lists
int<-rmvbin(75, margprob=eusi, bincorr=cori)
cont<-rmvbin(75, margprob=eusc, bincorr=corc)
```

```
#sum across rows by list
Sint<-rowSums (int, na.rm = FALSE, dims = 1)
Scontr<-rowSums (cont, na.rm = FALSE, dims = 1)
#estimate
Pi_s <- mean(Sint) - mean(Scontr)
return(Pi_s)

}
#This function finds the variance of n1 variances from n2 diff in means estimators using
the unequal and nonsymmetric distribution from above.
NRRICTUSV <- function(n1, n2) {
#create variance matrix with n1 rows.
Varmat <- matrix(nrow = n1, ncol=1)
        for (j in 1:n1) {
#create estimator matrix with n2 rows. Call function n2 times.
                MAT<- matrix(nrow = n2, ncol=1)
                        for (i in 1:n2) MAT[i,] <-NRRICTU()
                #find variance of each n2 sims.
        Varmat[j,] <-var(MAT)
        }
#find final variance.
        VarSim<-var(Varmat)
        return(VarSim)
}
#This function will determine variance for SSC – based on 200 simulations of 500
expected value estimators.
NRRSSC <- function() {

#based on sample smallest sample size (150), smallest pi_s (0.01), no correlation, largest
list size (5)

#no corr

cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
#input SSC distribution based on 5 item list.
ei<-c(1/2,1/2,1/2,1/2,1/2,.01)
#create list.
int<-rmvbin(150, margprob=ei, bincorr=cori)
#sum list.
Sint<-rowSums (int, na.rm = FALSE, dims = 1)
#expected value estimator.
Pi_s <- mean(Sint) - 2.5
return(Pi_s)
}
```

```
#This function will find the variance from n1 variances of n2 simulations.
NRRSSCV <- function(n1, n2) {
#Create variance matrix.
Varmat <- matrix(nrow = n1, ncol=1)
        for (j in 1:n1) {
        #create estimator variance.
                MAT<- matrix(nrow = n2, ncol=1)
#input n2 estimates.
                        for (i in 1:n2) MAT[i,] <-NRRSSC()
#find variance for n1 sims.
        Varmat[j,] <-var(MAT)
        }
#get overall variance.
        VarSim<-var(Varmat)
        return(VarSim)
}
```

R-Code: Determine Optimal Model for UQT compared to DQT

#Determines optimal model for UQT
#where Truth is defined as the proportion of participants with the sensitive attribute who respond truthfully, pins: proportion of participants with the non-sensitive attribute, p1 is the prob of selecting the sensitive question in sample 1, p2 is the probability of selecting the sensitive question in sample 2, n is the total number of participants.

```
UQT <- function(Truth,pins, pis, p1, p2, n) {
#Estimate proportion of those responding truthfully.
pis_hat <- Truth*pis
#estimate bias
bias <- pis_hat - pis
#calculate lambdas for each sample.
lambda1<-p1*pis_hat + (1-p1)*pins
lambda2<-p2*pis_hat + (1-p2)*pins
#create temp variables representing those not responding to sensitive question in each sample.
d1<-1-p2
d2<-1-p1
#allocates sample per Greenberg et al (1969) calculation.
r<- sqrt((lambda1*(1-lambda1)*d1^2)/(lambda2*(1-lambda2)*d2^2))

optn2<-round(n/(1+r),0)
optn1<- n-optn2
#calc variance components.
varc1<- (lambda1*(1-lambda1)*d1^2)/optn1
varc2<- (lambda2*(1-lambda2)*d2^2)/optn2
#denominator of variance.
d3 <-p1-p2
#calc variance
varpis <- (varc1 + varc2)/d3^2
#calc DQT variance.
vardq<-(pis_hat*(1-pis_hat))/n
#find variance for UQT (MSE) and DQ(RR)
MSE <- varpis + bias^2
rr <- vardq/MSE
#input results and return.
cell <- c(n, pis, pins, p1, p2, vardq, MSE, rr)
return(cell)
}

#Sim1 will run simulations taking rel ratio for all combinations of pi, n, p1 and pins
sim1<-	function(n) {
#defined all pi from study
sens		<-	c(0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45)
```

```
nonsens          <-    c(0.10, 0.20, 0.30) #non sens prev from study
psens           <-    c(0.90, 0.80,0.70, 0.60) #prob of selecting sens question
                mat<-matrix(nrow = 8, ncol=108)
#runs all combinations.
          l = 0
          for (i in 1:9) {
              for (j in 1:3) {

                    var1<-(12*i-11)+(j-1)*4
                                        var2<-(12*i-11)+(j-1)*4+3
                    for (k in var1:var2) mat[,k]<-UQT(1,nonsens[j],sens[i],psens[k-l*4],1-
psens[k-l*4],n)
                    l=l+1

                             }
                         }
                    return(mat)

}

#Function will call Sim1 above and write results to an EXCEL file.
runuqt <- function(n, outfile) {
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results"

          fmat<-matrix(nrow=8,ncol=108)
          fmat<- sim1(n)
          #return(fmat)
write.csv(fmat,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
}
```

RCODE for Forced Choice Technique compared to DQT

```
#Run FC method and outputs rel ratio as compared to DQT
#Truth: percent of sens subjects who report honestly
#pis: sens prevalence
#n1: sample size
FC <- function(Truth,pis, n1) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
#calcs lambda
lambda1<-(3/4)*pis_hat + 1/6
#calcs variance of FCT
varpis<- (lambda1*(1-lambda1))/(n1*(9/16))
#calcs var of DQT
vardq<-(pis_hat*(1-pis_hat))/n1
MSE <- varpis + bias^2
#rel ratio
rr <- vardq/MSE
#returns rel ratio
cell <- c(n1, pis,vardq, MSE, rr)
return(cell)
}

#Run all sens prev for specified n

simFC<-     function(n) {
#creates vector of sens prev rates from study
sens            <-   c(0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45)
#runs all combinations of prev rates and same sizes
                mat<-matrix(nrow = 5, ncol=9)
        l = 0
        for (i in 1:9) {

                mat[,i]<-FC(1,sens[i],n)
        l=l+1


                    }

                return(mat)

}
```

```
# Run sims for sample sizes
#Will call functions above and save rel ratios to excel spreadsheet.
runfc <- function(n, outfile) {
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\FCT"

	fmat<-matrix(nrow=5,ncol=9)
	fmat<- simFC(n)
	#return(fmat)
write.csv(fmat,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
}
```

R-Code: Determine Optimal Model ICT – 3, 4 and 5 Item lists.

```
#Uncorrelated, ICT – 3 Item

NRRICTEQ003 <- function(n,type, mp1, mp2, outfile) {
#where output will be save.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT3ItemUnCorr"
#define correlation.
cori<-cbind(c(1,0,0,0),c(0,1,0,0),c(0,0,1,0),c(0,0,0,1))
corc<-cbind(c(1,0,0),c(0,1,0),c(0,0,1))
#define non-sens prev rates for each group.
ei<-mp1
ec<-mp2
#performs 1000 sims
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                                #generate lists.
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)
                                #sum rows
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        #output estimate
                                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
 #find var
VarSim <-var(MAT)
#sensitive attribute
pis<-mp1[4]
#variance of DQT
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
#outputs rel ratio by dist type.
cell <- c(n,type, pis, vardq, VarSim, rr)
#puts in excel.
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)

}
```

#Uncorrelated: ICT-4 Item

```
NRRICTEQ004 <- function(n,type, mp1, mp2, outfile) {
#sets up excel sheet
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT4ItemUnCorr"
#covariance
cori<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0), c(0,0,0,0,1))
corc<-cbind(c(1,0,0,0),c(0,1,0,0),c(0,0,1,0),c(0,0,0,1))
#Non-sens distribution
ei<-mp1
ec<-mp2
#calculates lists, sums and estimates.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        #print(i)
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)

                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        #print(length(Sint))
                        #print(length(Scontr))

                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
 #variance
VarSim <-var(MAT)
#sensitive attribute being estimated.
pis<-mp1[5]
#DQT variance
vardq<-(pis*(1-pis))/(2*n)
#rel reli
rr <- vardq/VarSim
#outputs by non-sens dist type.
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#Uncorrelated ICT 5-Item

#Item list 5, Uncorrelated

```
NRRICTEQ005 <- function(n,type, mp1, mp2, outfile) {
#Excel spreadsheet for output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT5ItemUnCorr"
#correlation matrix
cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0),c(0,0,0,0,1))
#dist of sens and non-sens prevalent rates
ei<-mp1
ec<-mp2
#creates list, estimators
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        #print(i)
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)

                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        #print(length(Sint))
                        #print(length(Scontr))

                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
 #Calculates Rel Rel compared to DQT.
VarSim <-var(MAT)
#sens attribute being estimated
pis<-mp1[6]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
#outputs by distribution type.
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#Correlated ICT 3-Item

NRRICTEQC03 <- function(n,type, mp1, mp2, outfile) {
#Create excel doc for output
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT3ItemCorr"
#Define correlated matrix.
cori<-cbind(c(1,0,0,0),c(0,1,-.50,0),c(0,-.50,1,0),c(0,0,0,1))
corc<-cbind(c(1,0,0),c(0,1,-.50),c(0,-.50,1))
#Prevalent rates for intervention & control groups.
ei<-mp1
ec<-mp2
#Create lists, estimators
	MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
			int<-rmvbin(n, margprob=ei, bincorr=cori)
			contr<-rmvbin(n, margprob=ec, bincorr=corc)

			Sint<-rowSums (int, na.rm = FALSE, dims = 1)
			Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)

			MAT[i,] <- mean(Sint) - mean(Scontr)
				}
 #Output variance and compare to DQT.
VarSim <-var(MAT)
#sens prev rate being estimated
pis<-mp1[4]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
#output to excel.
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#Correlated ICT 4-Item

#Item 4 Correlated
#Correlate non-sensitive questions 1 and 2 and 3 and 4


NRRICTEQC04 <- function(n,type, mp1, mp2, outfile) {
#define excel document for output
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT4ItemCorr"
#define correlation matrix
cori<-cbind(c(1,-.50,0,0,0),c(-.50,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0), c(0,0,0,0,1))
corc<-cbind(c(1,-.50,0,0),c(-.50,1,0,0),c(0,0,1,-.50),c(0,0,-.50,1))
#prevalent rate distributions for cont and intervention groups.
ei<-mp1
ec<-mp2
#1000 sims that create lists, estimators and output to matrix.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        )
                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
 #variance
VarSim <-var(MAT)
#non sens being estimated.
pis<-mp1[5]
#DQT var
vardq<-(pis*(1-pis))/(2*n)
#rel ratio outputted
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#ICT - Unequal and Non-Symmetric, Correlated Type =2
#Correlate non-sensitive questions 2 and 4

NRRICTSYC04 <- function(n,type, mp1, mp2, outfile) {
#Create excel spreadsheet.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT4ItemCorr"
#define correlation matris
cori<-cbind(c(1,0,0,0,0),c(0,1,0,-.50,0),c(0,0,1,0,0),c(0,-.50,0,1,0), c(0,0,0,0,1))
corc<-cbind(c(1,0,0,0),c(0,1,0,-.50),c(0,0,1,0),c(0,-.50,0,1))
#prev rate distribution for controls/intervention.
ei<-mp1
ec<-mp2
#Will run 1000 sims, creating lists, estimators and finding variance.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
 #find variances and output rel rel to excel.
VarSim <-var(MAT)
#sens prevalent rate being estimated.
pis<-mp1[5]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#ICT - Unequal and Non-Symmetric, Correlated
#Correlate non-sensitive questions 3 and 4

NRRICTNSC04 <- function(n,type, mp1, mp2, outfile) {
#EXCEL file for output.
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT4ItemCorr"
#correlation
cori<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0), c(0,0,0,0,1))
corc<-cbind(c(1,0,0,0),c(0,1,0,0),c(0,0,1,-.50),c(0,0,-.50,1))
#distribution of prev rate for intervention and control.ei<-mp1
ec<-mp2
#1000 sims will generate list, estimates and take variance.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        int<-rmvbin(n, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n, margprob=ec, bincorr=corc)
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        MAT[i,] <- mean(Sint) - mean(Scontr)
        }
#find variances, calc rel rel and output
 VarSim <-var(MAT)
#sens prev rate estimated.
pis<-mp1[5]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#Correlated ICT 5-Item
#Item list 5, Correlated


NRRICTEQC05 <- function(n,type, mp1, mp2, outfile) {
#Excel for output.
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT5ItemCorr"
#define corr matrix
cori<-cbind(c(1,-.50,0,0,0,0),c(-.5,1,0,0,0,0),c(0,0,1,-.5,0,0),c(0,0,-.5,1,0,0),
c(0,0,0,0,1,0), c(0,0,0,0,0,1))
corc<-cbind(c(1,-.50,0,0,0),c(-.5,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0),c(0,0,0,0,1))
#prevalent rate dist for int and control samples.
ei<-mp1
ec<-mp2
#1000 sims, creates list and estimators. Variance is calculated.
	MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
			int<-rmvbin(n, margprob=ei, bincorr=cori)
			contr<-rmvbin(n, margprob=ec, bincorr=corc)
			Sint<-rowSums (int, na.rm = FALSE, dims = 1)
			Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
			MAT[i,] <- mean(Sint) - mean(Scontr)
				}
#Variances calculated and rel ratio calc and output.
VarSim <-var(MAT)
#sens rate estimated.
pis<-mp1[6]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}

#FOR BOTH UNEQUAL AND SYMM AS WELL AS UNEQUAL AND UNSYM,
CORRELATE NONSENSITIVE QUESTIONS 2 AND 5 AND 3 AND 4

NRRICTSYC05 <- function(n,type, mp1, mp2, outfile) {
#Excel for output.
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICT5ItemCorr"
#correlation matrix.
cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,-.50,0),c(0,0,1,-.50,0,0),c(0,0,-.50,1,0,0), c(0,-
.50,0,0,1,0), c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,-.50),c(0,0,1,-.50,0),c(0,0,-.50,1,0),c(0,-.50,0,0,1))
#prevalent rate distributions for control and intervention.
ei<-mp1
```

```
ec<-mp2
#1000 sims creating lists, estimators and taking variance.
	MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
			int<-rmvbin(n, margprob=ei, bincorr=cori)
			contr<-rmvbin(n, margprob=ec, bincorr=corc)
			Sint<-rowSums (int, na.rm = FALSE, dims = 1)
			Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
			MAT[i,] <- mean(Sint) - mean(Scontr)
			}
#Find variances and calc rel rel and output.
VarSim <-var(MAT)
#prev rate estimated.
pis<-mp1[6]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

R-Code: Optimal Model DICT, testing between, within and no correlation, 5-Item

#between correlation - 5 Item List

```
DICT5BT <- function(n, type, mp1, outfile) {
#Excel file for output.
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICTCORCK"
#correlation matrix for groups 1 and 2.
cors1<-cbind( c(	1,	0,	0,	0,	0,	0,	0.85,	0,	0,
	0,	0),
		c(	0,	1,	0,	0,	0,	0,	0,	0.85,
	0,	0,	0),
		c(	0,	0,	1,	0,	0,	0,	0,	0,
	0.85,	0,	0),
		c(	0,	0,	0,	1,	0,	0,	0,	0,
	0,	0.85,	0),
		c(	0,	0,	0,	0,	1,	0,	0,	0,
	0,	0,	0.85),
		c(	0,	0,	0,	0,	0,	1,	0,	0,
	0,	0,	0),
		c(	0.85,	0,	0,	0,	0,	0,	1,	0,
	0,	0,	0),
		c(	0,	0.85,	0,	0,	0,	0,	0,	1,
	0,	0,	0),
		c(	0,	0,	0.85,	0,	0,	0,	0,	0,
	1,	0,	0),
		c(	0,	0,	0,	0.85,	0,	0,	0,	0,
	0,	1,	0),
		c(	0,	0,	0,	0,	0.85,	0,	0,	0,
	0,	0,	1))
mp<-mp1
#1000 sims creating lists, estimators, and taking variance.
MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
			s1<-rmvbin(n, margprob=mp, bincorr=cors1)
			s2<-rmvbin(n, margprob=mp, bincorr=cors1)
			ints1<-s1[,1:6]
		conts1<-s1[,7:11]
			ints2<-s2[,1:6]
		conts2<-s2[,7:11]
			Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
			Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
			Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
			Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
			MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
```

```
                                        }
#variance.
VarSim <-var(MAT)
#sens estimated.
pis<-mp1[6]
#output.
cell <- c(type, pis, VarSim)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}

#Within correlation - 5 Item List

DICT5WI <- function(n, type, mp1, outfile) {
#Excel for output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICTCORCK"
#Correlation matrix for both samples.
cors1<-cbind(c(     1,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0),
                c(      0,      1,      0,      0,      -.50,   0,      0,      0,      0,
        0,      0),
                c(      0,      0,      1,      -.50,   0,      0,      0,      0,      0,
        0,      0),
                c(      0,      0,      -.50,   1,      0,      0,      0,      0,      0,
        0,      0),
                c(      0,      -.50,   0,      0,      1,      0,      0,      0,      0,
        0,      0),
                c(      0,      0,      0,      0,      0,      1,      0,      0,      0,
        0,      0),
                c(      0,      0,      0,      0,      0,      0,      1,      0,      0,
        0,      0),
                c(      0,      0,      0,      0,      0,      0,      0,      1,      0,
        0,      -.50),
                c(      0,      0,      0,      0,      0,      0,      0,      0,      1,
        -.50,   0),
                c(      0,      0,      0,      0,      0,      0,      0,      0,      -.50,
        1,      0),
                c(      0,      0,      0,      0,      0,      0,      0,      -.50,   0,
        0,      1))
#distribution of prev rates for each group.
mp<-mp1
#run 1000 sims, create lists, estimators and outputs to find variance.
MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        s1<-rmvbin(n, margprob=mp, bincorr=cors1)
                        s2<-rmvbin(n, margprob=mp, bincorr=cors1)
```

```
            ints1<-s1[,1:6]
        conts1<-s1[,7:11]
            ints2<-s2[,1:6]
        conts2<-s2[,7:11]
            Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
            Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
            Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
            Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
            MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
#Get variance.                    }
VarSim <-var(MAT)
#prev of sens estimated
pis<-mp1[6]
cell <- c(type, pis, VarSim)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#No correlation - 5 Item List

```
DICT5NO <- function(n, type, mp1, outfile) {
#excel for output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICTCORCK"
#correlation matrix.
cors1<-cbind(c(     1,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      0),
                c(  0,      1,      0,      0,      0,      0,      0,      0,      0,
        0,      0),
                c(  0,      0,      1,      0,      0,      0,      0,      0,      0,
        0,      0),
                c(  0,      0,      0,      1,      0,      0,      0,      0,      0,
        0,      0),
                c(  0,      0,      0,      0,      1,      0,      0,      0,      0,
        0,      0),
                c(  0,      0,      0,      0,      0,      1,      0,      0,      0,
        0,      0),
                c(  0,      0,      0,      0,      0,      0,      1,      0,      0,
        0,      0),
                c(  0,      0,      0,      0,      0,      0,      0,      1,      0,
        0,      0),
                c(  0,      0,      0,      0,      0,      0,      0,      0,      1,
        0,      0),
                c(  0,      0,      0,      0,      0,      0,      0,      0,      0,
        1,      0),
                c(  0,      0,      0,      0,      0,      0,      0,      0,      0,
        0,      1))
#prev rate distribution for both samples.
mp<-mp1
#Runs 1000 sims, creating lists, sums, calc estimators and finding variance.
MAT <- matrix(nrow = 1000, ncol=1)
        for (i in 1:1000) {
                s1<-rmvbin(n, margprob=mp, bincorr=cors1)
                s2<-rmvbin(n, margprob=mp, bincorr=cors1)
                ints1<-s1[,1:6]
            conts1<-s1[,7:11]
                ints2<-s2[,1:6]
            conts2<-s2[,7:11]
                Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
                MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
                        }
```

```
#output variances.
VarSim <-var(MAT)
pis<-mp1[6]
cell <- c(type, pis, VarSim)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

R-Code: Optimal Model DICT (Between Correlation), testing between equal, symmetric & not equal, not symmetric & not equal. Item list size 3, 4 and 5

```
#Item list size 3

#between correlation - 3 Item List

DICT3BT <- function(n, type, mp1, outfile) {
#excel for output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICT3ItemBet"
#correlation matrix.
cors1<-cbind( c(1,   0,    0,    0,    .85,  0,    0),
              c(0,   1,    0,    0,    0,    .85,  0),
              c(0,   0,    1,    0,    0,    0,    .85),
              c(0,   0,    0,    1,    0,    0,    0),
              c(.85,0,     0,    0,    1,    0,    0),
              c(0,   .85,  0,    0,    0,    1,    0),
              c(0,   0,    .85,  0,    0,    0,    1))
#prev rate distr for both samples.
mp<-mp1
#1000 sims, create lists, sums rows, calcs estimators.
MAT <- matrix(nrow = 1000, ncol=1)
              for (i in 1:1000) {
                    s1<-rmvbin(n, margprob=mp, bincorr=cors1)
                    s2<-rmvbin(n, margprob=mp, bincorr=cors1)
                    ints1<-s1[,1:4]
                 conts1<-s1[,5:7]
                    ints2<-s2[,1:4]
                 conts2<-s2[,5:7]
                    Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                    Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                    Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                    Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
                    MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
                          }
#variances calculated and rel ratio compared to DQT and output.
VarSim <-var(MAT)
#sens prev rate.
pis<-mp1[4]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#DICT Item list size 4

#between correlation - 4 Item List

DICT4BT <- function(n, type, mp1, outfile) {
#EXCEL for output.
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICT4ItemBet"
#correlation matrix for both sample.
cors1<-cbind( c(1,     0,      0,      0,      0,      .85,    0,      0,      0),
                c(0,     1,      0,      0,      0,      0,      0.85,   0,      0),
                c(0,     0,      1,      0,      0,      0,      0,      0.85,   0),
                c(0,     0,      0,      1,      0,      0,      0,      0,      0.85),
                c(0,     0,      0,      0,      1,      0,      0,      0,      0),
                c(0.85,  0,      0,      0,      0,      1,      0,      0,      0),
                c(0,     0.85,   0,      0,      0,      0,      1,      0,      0),
                c(0,     0,      0.85,   0,      0,      0,      0,      1,      0),
                c(0,     0,      0,      0.85,   0,      0,      0,      0,      1))
#prevalent rate distr for both samples.
mp<-mp1
#1000 sims, creates lists, sums rows, calcs estimators.
MAT <- matrix(nrow = 1000, ncol=1)
        for (i in 1:1000) {
                s1<-rmvbin(n, margprob=mp, bincorr=cors1)
                s2<-rmvbin(n, margprob=mp, bincorr=cors1)
                ints1<-s1[,1:5]
            conts1<-s1[,6:9]
                ints2<-s2[,1:5]
            conts2<-s2[,6:9]
                Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
                MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
                        }
#Calcs variance, outputs rel ratio for prev rate being estimated.
VarSim <-var(MAT)
pis<-mp1[5]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#DICT, Item list size 5

#between correlation - 5 Item List

DICT5BT <- function(n, type, mp1, outfile) {
#Excel for output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\DICT5ItemBet"
#correlation matrix.
cors1<-cbind( c(    1,      0,      0,      0,      0,      0,      0.85,   0,      0,
        0,      0),
                    c(      0,      1,      0,      0,      0,      0,      0,      0.85,
        0,      0,      0),
                    c(      0,      0,      1,      0,      0,      0,      0,      0,
        0.85,   0,      0),
                    c(      0,      0,      0,      1,      0,      0,      0,      0,
        0,      0.85,   0),
                    c(      0,      0,      0,      0,      1,      0,      0,      0,
        0,      0,      0.85),
                    c(      0,      0,      0,      0,      0,      1,      0,      0,
        0,      0,      0),
                    c(      0.85,   0,      0,      0,      0,      0,      1,      0,
        0,      0,      0),
                    c(      0,      0.85,   0,      0,      0,      0,      0,      1,
        0,      0,      0),
                    c(      0,      0,      0.85,   0,      0,      0,      0,      0,
        1,      0,      0),
                    c(      0,      0,      0,      0.85,   0,      0,      0,      0,
        0,      1,      0),
                    c(      0,      0,      0,      0,      0.85,   0,      0,      0,
        0,      0,      1))
mp<-mp1 #prevalent rate distr for both samples.
#1000 sims. Creates list, sums rows, calcs estimator.
MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                    s1<-rmvbin(n, margprob=mp, bincorr=cors1)
                    s2<-rmvbin(n, margprob=mp, bincorr=cors1)
                    ints1<-s1[,1:6]
                conts1<-s1[,7:11]
                    ints2<-s2[,1:6]
                conts2<-s2[,7:11]
                    Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                    Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                    Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                    Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
                    MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2}                  }
```

```
#Variances from each technique compared and outputted by sens prev rate.
VarSim <-var(MAT)
pis<-mp1[6]
vardq<-(pis*(1-pis))/(2*n)
rr <- vardq/VarSim
cell <- c(n,type, pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

R-CODE: SSC, Finding optimal model, List sizes 3, 4 and 5

```
#SSC, Optimal Model List Size 5
#Not Correlated
NRRSSC05 <- function(n, mp, outfile) {
#excel spreadsheet for ouput.
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC5ItemUnCorr"
#corr Q1 AND Q2, Q3 AND Q4
cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,0,0),c(0,0,1,0,0,0),c(0,0,0,1,0,0), c(0,0,0,0,1,0),
c(0,0,0,0,0,1))
#prevalent rate dist.
ei<-mp
#1000 sims, create lists, sums rows, calcs estimator.
	MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
		int<-rmvbin(n, margprob=ei, bincorr=cori)
		Sint<-rowSums (int, na.rm = FALSE, dims = 1)
		MAT[i,] <- mean(Sint) - 2.5
						}
#calcs variances for each technique and outputs by sens prev rate.
VarSim <-var(MAT)
pis<-mp[6]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#SSC, Optimal Model List Size 4

```
NRRSSC04 <- function(n, mp, outfile) {
#Excel for output.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC4ItemUnCorr"
#correlation matrix.
cori<-cbind(c(1,0,0,0,0),c(0,1,0,0,0),c(0,0,1,0,0),c(0,0,0,1,0), c(0,0,0,0,1))
#prev rate distr
ei<-mp
#1000 sims creating lists, summing rows, calc estimators.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 2
                                        }
#variances calc, rel rel outputted by sens prev rate estimated.
VarSim <-var(MAT)
pis<-mp[5]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
SETUPSSC04("OMSSC4_0")

#SSC, Optimal Model List Size 3, No correlation.

NRRSSC03 <- function(n, mp, outfile) {
#EXCEL output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC3ItemUnCorr"
#No Correlation
cori<-cbind(c(1,0,0,0),c(0,1,0,0),c(0,0,1,0),c(0,0,0,1))
#prev rate dist
ei<-mp
#1000 sims, creating lists, sum rows, calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 1.5
                                        }
#variance from each technique calc, rel rel calc and outputted by prevalent rate estimated.
VarSim <-var(MAT)
pis<-mp[4]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#Correlated

```
#SSC, Optimal Model List Size 5
#Correlated
NRRSSCC5 <- function(n, mp, outfile) {
#Excel output.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC5ItemCorr"
#corr Q1 AND Q2, Q3 AND Q4
cori<-cbind(c(1,-.50,0,0,0,0),c(-.5,1,0,0,0,0),c(0,0,1,-.5,0,0),c(0,0,-.5,1,0,0),
c(0,0,0,0,1,0), c(0,0,0,0,0,1))
#prev rate distribution.
ei<-mp
#1000 sims, creating list, sum rows, calc estimators.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 2.5
                                        }
#calc variances, calc rel rel, output by sens prev rate estimated.
VarSim <-var(MAT)
pis<-mp[6]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
SETUPSSCC5("OMSSC5_C")
#SSC, Optimal Model List Size 4
NRRSSCC4 <- function(n, mp, outfile) {
#excel output.
area    <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC4ItemCorr"
#CORRELATE Q1 AND Q2, Q3 AND Q4
cori<-cbind(c(1,-.50,0,0,0),c(-.50,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0), c(0,0,0,0,1))
#prev rate distribution.
ei<-mp
#1000 sims, creates list, sums row, calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 2
                                }
#calcs variance, rel rel and outputs by sens prev rate estimated.
VarSim <-var(MAT)
pis<-mp[5]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#SSC, Optimal Model List Size 3

```
NRRSSCC3 <- function(n, mp, outfile) {
#EXCEL output.
area    <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\SSC3ItemCorr"
#Correlate q2 and q3
cori<-cbind(c(1,0,0,0),c(0,1,-.50,0),c(0,-.50,1,0),c(0,0,0,1))
#prev rate distribution.
ei<-mp
#1000 sims, creates lists, sums rows, calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 1.5
                                        }
#cals variances, rel rel, outputs by sens prev rate estimated.
VarSim <-var(MAT)
pis<-mp[4]
vardq<-(pis*(1-pis))/n
rr <- vardq/VarSim
cell <- c(n,pis, vardq, VarSim, rr)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

R-Code: Noncompliance, UQT Optimal, UQT Practical, FCT vs. ICT optimal 3-Item, 4-Item, 5-Item lists.

```
#Optimal Model for UQT based on efficiency test is: p1=.90, p2=.10, pi_ns = .10 - will set parameters to these values

#Optimal Models:
#UQT - p1=.9, pins =.10 (Optimal)
#UQT - p1=.6, pins =.30 (Practical)
#ICT - Item 3 (corr, nonsym)
        #Item 4 (corr, equal)
        #Item 5 (corr, symm)
#DICT- #Item 3 (corr between, equal)
        #Item 4 (corr between, nonsym)
        #Item 5 (corr between, sym)
#SSC - All models (corr)

#Start with optimal model UQT

UQTOP <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis

lambda1<-.90*pis_hat + (1-.90)*.10
lambda2<-.10*pis_hat + (1-.10)*.10
#allocates sample sizes per Greenberg et al (1969)
r<- sqrt((lambda1*(1-lambda1)*.9^2)/(lambda2*(1-lambda2)*.1^2))
#sample sizes calculated.
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#variance components calc.
varc1<- (lambda1*(1-lambda1)*.90^2)/optn1
varc2<- (lambda2*(1-lambda2)*.10^2)/optn2
#variance cal.
varpis <- (varc1 + varc2)/.80^2
#MSE calc.
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}
```

```
#Practical UQT Model
#Truth – percent of those w/ sens attribute who respond truthfully.
UQTPRAC <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-.60*pis_hat + (1-.60)*.30
lambda2<-.40*pis_hat + (1-.40)*.30
#allocate sample.
r<- sqrt((lambda1*(1-lambda1)*.6^2)/(lambda2*(1-lambda2)*.4^2))
#create n for each sample.
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#calc variance components.
varc1<- (lambda1*(1-lambda1)*.60^2)/optn1
varc2<- (lambda2*(1-lambda2)*.40^2)/optn2
#calc variance and MSE.
varpis <- (varc1 + varc2)/.20^2
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}


#Run FC method
#Truth – percent of those w/ sens attribute who respond truthfully.

FC <- function(Truth,pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
#variance component.
lambda1<-(3/4)*pis_hat + 1/6
#variance.
varpis<- (lambda1*(1-lambda1))/(n*(9/16))
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}
```

```
#3ICT - Correlated with nonsym - optimal model

NRRICTEQC03 <- function(Truth,pis, n, outfile) {
#Excel output.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cori<-cbind(c(1,0,0,0),c(0,1,-.50,0),c(0,-.50,1,0),c(0,0,0,1))
corc<-cbind(c(1,0,0),c(0,1,-.50),c(0,-.50,1))
#simulate bias data
#prev rate dist by int and control group.
ei<-c(1/4,2/3,2/3,pis_hat)
ec<-c(1/4,2/3,2/3)
#1000 sims, create lists, sums rows and calc estimates.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        int<-rmvbin(n/2, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n/2, margprob=ec, bincorr=corc)
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
#calc bias and MSE – output.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}

#Equal 4 –item list.
NRRICTEQC04 <- function(Truth,pis, n, outfile) {
#Define output.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cori<-cbind(c(1,-.50,0,0,0),c(-.50,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0), c(0,0,0,0,1))
corc<-cbind(c(1,-.50,0,0),c(-.50,1,0,0),c(0,0,1,-.50),c(0,0,-.50,1))
#prev rate distr for each sample.
ei<-c(2/3,2/3,2/3,2/3,pis_hat)
ec<-c(2/3,2/3,2/3,2/3)
#1000 sims. Creates list, sums rows, calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
```

```
                    int<-rmvbin(n/2, margprob=ei, bincorr=cori)
                    contr<-rmvbin(n/2, margprob=ec, bincorr=corc)
                    Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                    Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                    MAT[i,] <- mean(Sint) - mean(Scontr)
                        }
#Find bias and output MSE.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#ICT 5-Item
NRRICTSYC05 <- function(Truth,pis, n, outfile) {
#excel output.
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cori<-cbind(c(1,0,0,0,0,0),c(0,1,0,0,-.50,0),c(0,0,1,-.50,0,0),c(0,0,-.50,1,0,0), c(0,-
.50,0,0,1,0), c(0,0,0,0,0,1))
corc<-cbind(c(1,0,0,0,0),c(0,1,0,0,-.50),c(0,0,1,-.50,0),c(0,0,-.50,1,0),c(0,-.50,0,0,1))
#prev distr.
ei<-c(1/6,2/6,3/6,4/6,5/6,pis_hat)
ec<-c(1/6,2/6,3/6,4/6,5/6)
#1000 sims, creates list, sums rows and calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                        int<-rmvbin(n/2, margprob=ei, bincorr=cori)
                        contr<-rmvbin(n/2, margprob=ec, bincorr=corc)
                        Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                        Scontr<-rowSums (contr, na.rm = FALSE, dims = 1)
                        MAT[i,] <- mean(Sint) - mean(Scontr)
                                }
#cal MSE and output.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#Simulate and save MSE
#Runs each of the sims above by sample size and calculates ratio MSE between all RR
and NRR techniques.

NCSIM1 <- function(Truth,pis, n)  {
        UQTOMSE <-UQTOP(Truth, pis, n)
        UQTPMSE <- UQTPRAC(Truth, pis, n)
        FCMSE <-FC(Truth, pis, n)
        ICT3 <- NRRICTEQC03(Truth, pis, n, "ICT3")
        ICT4 <- NRRICTEQC04(Truth, pis, n, "ICT4")
        ICT5 <-NRRICTSYC05(Truth, pis, n, "ICT5")
#output and calc MSEs
cell <- c(Truth, pis, n, UQTOMSE, UQTPMSE, FCMSE, ICT3, ICT4, ICT5,
UQTOMSE/ICT3, UQTOMSE/ICT4, UQTOMSE/ICT5,
                        UQTPMSE/ICT3, UQTPMSE/ICT4, UQTPMSE/ICT5,
                        FCMSE/ICT3, FCMSE/ICT4, FCMSE/ICT5)
return(cell)
}

#Run sim above for each non-compliance rate defined in study

NCSIM2<-      function(n) {
#define sens prev rates and non-compl rates for study.
sens          <-    c(0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45)
noncomp           <-    c(0.90, 0.80,0.70, 0.60)
#create matrix for output and run all combinations using the R-functions defined above.
             matr<-matrix(nrow = 18, ncol=36)
        l = 0
        for (i in 1:9) { var1<-i +(i-1)*3
                 var2<-i +(i-1)*3+ 3
                                for (k in var1:var2) matr[,k]<-NCSIM1(noncomp[k-
l*4],sens[i],n)
                l=l+1}
                return(matr)
               }

#run simulations and output to excel spreadsheet.
runncsim <- function(n, outfile) {
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
        fmat<-matrix(nrow=18,ncol=36)
        fmat<- NCSIM2(n)
        #return(fmat)
print(fmat)
write.csv(fmat,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
}
```

R-Code: Noncompliance, UQT Optimal, UQT Practical, FCT vs. DICT optimal 3-Item, 4-Item, 5-Item lists.

```
#Optimal Model for UQT based on efficiency test is: p1=.90, p2=.10, pi_ns = .10 - will set parameters to these values
#Optimal Models:
#UQT - p1=.9, pins =.10 (Optimal)
#UQT - p1=.6, pins =.30 (Practical)
#DICT- #Item 3 (corr between, equal)
        #Item 4 (corr between, nonsym)
        #Item 5 (corr between, sym)

#Start with optimal model UQT – this is same function defined previously.

UQTOP <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-.90*pis_hat + (1-.90)*.10
lambda2<-.10*pis_hat + (1-.10)*.10
#allocate sample
r<- sqrt((lambda1*(1-lambda1)*.9^2)/(lambda2*(1-lambda2)*.1^2))
#allocate sample
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#variance components.
varc1<- (lambda1*(1-lambda1)*.90^2)/optn1
varc2<- (lambda2*(1-lambda2)*.10^2)/optn2
#calc variance and MSE and output.
varpis <- (varc1 + varc2)/.80^2
MSE <- varpis + bias^2
cell <- c(MSE)
return(cell)
}

#Practical UQT Model

UQTPRAC <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-.60*pis_hat + (1-.60)*.30
lambda2<-.40*pis_hat + (1-.40)*.30
#allocate sample.
r<- sqrt((lambda1*(1-lambda1)*.6^2)/(lambda2*(1-lambda2)*.4^2))
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#variance components.
```

```
varc1<- (lambda1*(1-lambda1)*.60^2)/optn1
varc2<- (lambda2*(1-lambda2)*.40^2)/optn2
#calc variance and MSE and output.
varpis <- (varc1 + varc2)/.20^2
MSE <- varpis + bias^2
cell <- c(MSE)
return(cell)
}

#Run FC method
FC <- function(Truth,pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-(3/4)*pis_hat + 1/6
#varaince.
varpis<- (lambda1*(1-lambda1))/(n*(9/16))
#MSE and output.
MSE <- varpis + bias^2
cell <- c(MSE)
return(cell)
}
```

```
#3DICT - Correlated with nonsym - optimal model

DICT3BT <- function(Truth, pis, n, outfile) {
#output excel.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cors1<-cbind( c(1,    0,     0,     0,     .85,   0,     0),
              c(0,    1,     0,     0,     0,     .85,   0),
              c(0,    0,     1,     0,     0,     0,     .85),
              c(0,    0,     0,     1,     0,     0,     0),
              c(.85,0,       0,     0,     1,     0,     0),
              c(0,    .85,   0,     0,     0,     1,     0),
              c(0,    0,     .85,   0,     0,     0,     1))
#equal
mp<-c(2/3,2/3,2/3, pis_hat ,2/3,2/3,2/3)
#1000 sims. Creates list. Sums rows and calc estimator.
MAT <- matrix(nrow = 1000, ncol=1)
              for (i in 1:1000) {
                     s1<-rmvbin(n/2, margprob=mp, bincorr=cors1)
                     s2<-rmvbin(n/2, margprob=mp, bincorr=cors1)
                     ints1<-s1[,1:4]
                     conts1<-s1[,5:7]
                     ints2<-s2[,1:4]
                     conts2<-s2[,5:7]
                     Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                     Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                     Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                     Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
                     MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2                       }
#calc bias, MSE and output.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)}
```

```
#between correlation - 4 Item List

DICT4BT <- function(Truth,pis, n, outfile) {
#output to excel
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cors1<-cbind( c(1,    0,      0,      0,      0,      .85,    0,      0,      0),
              c(0,    1,      0,      0,      0,      0,      0.85,   0,      0),
              c(0,    0,      1,      0,      0,      0,      0,      0.85,   0),
              c(0,    0,      0,      1,      0,      0,      0,      0,      0.85),
              c(0,    0,      0,      0,      1,      0,      0,      0,      0),
              c(0.85, 0,      0,      0,      0,      1,      0,      0,      0),
              c(0,    0.85,   0,      0,      0,      0,      1,      0,      0),
              c(0,    0,      0.85,   0,      0,      0,      0,      1,      0),
              c(0,    0,      0,      0.85,   0,      0,      0,      0,      1))
#nonsym
mp<-c(1/6,3/6,4/6,4/6, pis_hat,1/6,3/6,4/6,4/6)
#1000 sims. Creates list, sums rows, calcs estimator.
MAT <- matrix(nrow = 1000, ncol=1)
        for (i in 1:1000) {
                s1<-rmvbin(n/2, margprob=mp, bincorr=cors1)
                s2<-rmvbin(n/2, margprob=mp, bincorr=cors1)
                ints1<-s1[,1:5]
            conts1<-s1[,6:9]
                ints2<-s2[,1:5]
            conts2<-s2[,6:9]
                Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
                Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
                Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
                Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
        MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
                        }
#calc bias and output MSE.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#between correlation - 5 Item List

DICT5BT <- function(Truth,pis, n, outfile) {
#Excel for output.
area	<-	"C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
cors1<-cbind( c(	1,	0,	0,	0,	0,	0,	0.85,	0,	0,
	0,	0),
		c(	0,	1,	0,	0,	0,	0,	0,	0.85,
	0,	0,	0),
		c(	0,	0,	1,	0,	0,	0,	0,	0,
	0.85,	0,	0),
		c(	0,	0,	0,	1,	0,	0,	0,	0,
	0,	0.85,	0),
		c(	0,	0,	0,	0,	1,	0,	0,	0,
	0,	0,	0.85),
		c(	0,	0,	0,	0,	0,	1,	0,	0,
	0,	0,	0),
		c(	0.85,	0,	0,	0,	0,	0,	1,	0,
	0,	0,	0),
		c(	0,	0.85,	0,	0,	0,	0,	0,	1,
	0,	0,	0),
		c(	0,	0,	0.85,	0,	0,	0,	0,	0,
	1,	0,	0),
		c(	0,	0,	0,	0.85,	0,	0,	0,	0,
	0,	1,	0),
		c(	0,	0,	0,	0,	0.85,	0,	0,	0,
	0,	0,	1))
#symmetric
mp<-c(1/6,2/6,3/6,4/6,5/6, pis_hat,1/6,2/6,3/6,4/6,5/6)
#100 sims, creates list, sums rows, calcs estimator.
MAT <- matrix(nrow = 1000, ncol=1)
		for (i in 1:1000) {
			s1<-rmvbin(n/2, margprob=mp, bincorr=cors1)
		s2<-rmvbin(n/2, margprob=mp, bincorr=cors1)
			ints1<-s1[,1:6]
		    conts1<-s1[,7:11]
			ints2<-s2[,1:6]
		    conts2<-s2[,7:11]
			Sints1<-rowSums (ints1, na.rm = FALSE, dims = 1)
			Sconts1<-rowSums (conts1, na.rm = FALSE, dims = 1)
			Sints2<-rowSums (ints2, na.rm = FALSE, dims = 1)
			Sconts2<-rowSums (conts2, na.rm = FALSE, dims = 1)
```

```
                              MAT[i,] <- ((mean(Sints1) - mean(Sconts2)) + (mean(Sints2) -
mean(Sconts1)))/2
                                        }
#Calc bias and output MSE.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}

#Simulate and save MSE

NCSIM1 <- function(Truth,pis, n)  {
#run all sims.
          UQTOMSE <-UQTOP(Truth, pis, n)
          UQTPMSE <- UQTPRAC(Truth, pis, n)
          FCMSE <-FC(Truth, pis, n)
          DICT3 <- DICT3BT(Truth, pis, n, "DICT3")
          DICT4 <- DICT4BT(Truth, pis, n, "DICT4")
          DICT5 <- DICT5BT(Truth, pis, n, "DICT5")
#calc mse between RR and NRR techniques.
cell <- c(Truth, pis, n, UQTOMSE, UQTPMSE, FCMSE, DICT3, DICT4, DICT5,
UQTOMSE/DICT3, UQTOMSE/DICT4, UQTOMSE/DICT5,
                              UQTPMSE/DICT3, UQTPMSE/DICT4, UQTPMSE/DICT5,
                              FCMSE/DICT3, FCMSE/DICT4, FCMSE/DICT5)
return(cell)
}

#Run by non-compliance rate
#Runs all simulations in prevalent rate, non-compliance rate defined in study – for DICT.
NCSIM2<-      function(n) {
sens             <-    c(0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45)
noncomp            <-    c(0.90, 0.80,0.70, 0.60)
                matr<-matrix(nrow = 18, ncol=36)
          l = 0
          for (i in 1:9) {
                                        var1<-i +(i-1)*3
                    var2<-i +(i-1)*3+ 3
                                       for (k in var1:var2) matr[,k]<-NCSIM1(noncomp[k-
l*4],sens[i],n)
                    l=l+1
                       }
                    return(matr)
}
```

```
#runs all sims above and saves to excel file.
runncsim <- function(n, outfile) {
area    <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSEFN"
        fmat<-matrix(nrow=18,ncol=36)
        fmat<- NCSIM2(n)
        #return(fmat)
write.csv(fmat,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
}
```

R-Code: Noncompliance, UQT Optimal, UQT Practical, FCT vs. SSC optimal 3-Item, 4-Item, 5-Item lists.

```
#Optimal Model for UQT based on efficiency test is: p1=.90, p2=.10, pi_ns = .10 - will set parameters to these values
#Optimal Models:
#UQT - p1=.9, pins =.10 (Optimal)
#UQT - p1=.6, pins =.30 (Practical)
#ICT - Item 3 (corr, nonsym)
        #Item 4 (corr, equal)
        #Item 5 (corr, symm)
#DICT- #Item 3 (corr between, equal)
        #Item 4 (corr between, nonsym)
        #Item 5 (corr between, sym)
#SSC - All models (corr)
#Start with optimal model UQT

UQTOP <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-.90*pis_hat + (1-.90)*.10
lambda2<-.10*pis_hat + (1-.10)*.10
#allocate samples
r<- sqrt((lambda1*(1-lambda1)*.9^2)/(lambda2*(1-lambda2)*.1^2))
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#variance components.
varc1<- (lambda1*(1-lambda1)*.90^2)/optn1
varc2<- (lambda2*(1-lambda2)*.10^2)/optn2
#variance.
varpis <- (varc1 + varc2)/.80^2
#MSE
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}
```

```
#Practical UQT Model

UQTPRAC <- function(Truth, pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-.60*pis_hat + (1-.60)*.30
lambda2<-.40*pis_hat + (1-.40)*.30
#allocate sample
r<- sqrt((lambda1*(1-lambda1)*.6^2)/(lambda2*(1-lambda2)*.4^2))
optn2<-round(n/(1+r),0)
optn1<- n-optn2
#variance components.
varc1<- (lambda1*(1-lambda1)*.60^2)/optn1
varc2<- (lambda2*(1-lambda2)*.40^2)/optn2
#variance.
varpis <- (varc1 + varc2)/.20^2
#MSE
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}

#Run FC method

FC <- function(Truth,pis, n) {
pis_hat <- Truth*pis
bias <- pis_hat - pis
lambda1<-(3/4)*pis_hat + 1/6
#variance.
varpis<- (lambda1*(1-lambda1))/(n*(9/16))
#MSE.
MSE <- varpis + bias^2
#output.
cell <- c(MSE)
return(cell)
}
```

#SSC, Optimal Model List Size 3

```
NRRSSCC3 <- function(Truth,pis, n, outfile) {
#EXCEL output.
#area   <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
#Correlate q2 and q3
cori<-cbind(c(1,0,0,0),c(0,1,-.50,0),c(0,-.50,1,0),c(0,0,0,1))
#simulate bias data
ei<-c(1/2,1/2,1/2,pis_hat)
#1000 sims, creates list, sums rows and calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 1.5
                                        }
#calc bias, MSE and output.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
#write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

#SSC, Optimal Model List Size 4

```
NRRSSCC4 <- function(Truth,pis, n, outfile) {
#define output.
area     <-      "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
#CORRELATE Q1 AND Q2, Q3 AND Q4
cori<-cbind(c(1,-.50,0,0,0),c(-.50,1,0,0,0),c(0,0,1,-.50,0),c(0,0,-.50,1,0), c(0,0,0,0,1))
#prevalence rate dist.
ei<-c(1/2,1/2,1/2,1/2,pis_hat)
#1000 sims, creates list, sums rows, calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 2
                                        }
#calcs bias, MSE and outputs.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}
```

```
#SSC, Optimal Model List Size 5
#Correlated
NRRSSCC5 <- function(Truth,pis, n, outfile)  {
#output excel.
area     <-        "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
#Calculate the pis in presence of non-compliance
pis_hat <- Truth*pis
#Correlation matrix
#corr Q1 AND Q2, Q3 AND Q4
cori<-cbind(c(1,-.50,0,0,0,0),c(-.5,1,0,0,0,0),c(0,0,1,-.5,0,0),c(0,0,-.5,1,0,0),
c(0,0,0,0,1,0), c(0,0,0,0,0,1))
#prev rate distribution.
ei<-c(1/2,1/2,1/2,1/2,1/2,pis_hat)
#1000 sims, calcs list, sums rows and calcs estimator.
        MAT <- matrix(nrow = 1000, ncol=1)
                for (i in 1:1000) {
                int<-rmvbin(n, margprob=ei, bincorr=cori)
                Sint<-rowSums (int, na.rm = FALSE, dims = 1)
                MAT[i,] <- mean(Sint) - 2.5
                                }
#calc bias, MSE and output.
Bias <- mean(MAT)- pis
MSE <- var(MAT) + Bias^2
cell <- c(MSE)
write.csv(MAT,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
return(cell)
}

#Simulate and save MSE

NCSIM1 <- function(Truth,pis, n)  {
#sims and outputs all MSE for SSC functions.
        UQTOMSE <-UQTOP(Truth, pis, n)
        UQTPMSE <- UQTPRAC(Truth, pis, n)
        FCMSE <-FC(Truth, pis, n)
        SSC3 <- NRRSSCC3(Truth, pis, n, "SSC3")
        SSC4 <- NRRSSCC4(Truth, pis, n, "SSC4")
        SSC5 <-NRRSSCC5(Truth, pis, n, "SSC5")
cell <- c(Truth, pis, n, UQTOMSE, UQTPMSE, FCMSE, SSC3, SSC4, SSC5,
UQTOMSE/SSC3, UQTOMSE/SSC4, UQTOMSE/SSC5,
                        UQTPMSE/SSC3, UQTPMSE/SSC4, UQTPMSE/SSC5,
                        FCMSE/SSC3, FCMSE/SSC4, FCMSE/SSC5)
return(cell)
}
```

```
#Run by non-compliance rate and sens prev rate combination for SSC.

NCSIM2<-     function(n) {
sens          <-    c(0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45)
noncomp           <-    c(0.90, 0.80,0.70, 0.60)
                matr<-matrix(nrow = 18, ncol=36)
        l = 0
        for (i in 1:9) {
                                var1<-i +(i-1)*3
                   var2<-i +(i-1)*3+ 3

                                for (k in var1:var2) matr[,k]<-NCSIM1(noncomp[k-
l*4],sens[i],n)
                l=l+1
                  }
                return(matr)
}

#runs all sims for SSC and saves output to excel.
runncsim <- function(n, outfile) {
area     <-       "C:\\D_DRIVE_Backup\\Dissertation\\Results\\ICTMSE"
        fmat<-matrix(nrow=18,ncol=36)
        fmat<- NCSIM2(n)
        #return(fmat)
print(fmat)
write.csv(fmat,paste(area,paste(outfile,"csv",sep="."),sep="\\"),row.names=F)
}
```