February 2018

Strategies to Adjust for Response Bias in Clinical Trials: A Simulation Study

Victoria R. Swaidan
University of South Florida, vswaidan@gmail.com

Strategies to Adjust for Response Bias in Clinical Trials:

A Simulation Study

by

Victoria R. Swaidan

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Public Health
with a concentration in Biostatistics
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida

Major Professor: Wei Wang, Ph.D.
Yangxin Huang, Ph.D.
Henian Chen, M.D., Ph.D.

Date of Approval:
February 5, 2018

Keywords: survey research, self-report bias, bias correction,
intervention studies, treatment effect, estimation, reliability, internal validity

Dedication

I dedicate this work to my loving and devoted family, without whom I would not be where I am today. To my sweet, beautiful daughter Aliyah, whose joy and laughter helped me get through so many long days/nights of studying and working on this project. To my husband Floyd, who kept me strong, helped me push through the trying times, and always kept me focused and motivated. And to my mother Peggy, who was full of encouragement and support, always helped me keep the end in sight, and devotingly sacrificed her working career to help make our house a home during these extraordinarily busy times. There are no words or gestures that could ever fully express the gratitude I feel for what each you have done to help me along this incredible journey. I truly appreciate each of you and thank you all from the bottom of my heart.

Acknowledgments

I express my deepest gratitude to my major professor Dr. Wei Wang for supporting me in the extention of his work in this area and for making this project possible. I am truly grateful for the immense amount of guidance, training, and mentorship that he provided throughout the entire thesis process. And to my committee members, Drs. Yangxin Huang and Henian Chen for the time and effort dedicated to the development and completion of this project. Without the guidance of this incredeible committee, this project would not be near the level of rigor that it is, and for that I am truly thankful.

A special thank you to the faculty of the Department of Epidemiology and Biostatistics in the College of Public Health at the University of South Florida for providing an incredible curriculum and learning environment that allowed me to pursue this degree.

Lastly, I express my sincere appreciation to Dr. Alison Salloum for the many years of research experience that she has provided, allowing me to grow and develop as a researcher under her mentorship. The experience I gained working on her clinical trials piqued my interest in statistics, inevitably bringing me into this program.

Table of Contents

List of Tables iii

List of Figures iv

Abstract v

Chapter 1: Overview of Bias 1
- Bias in the Design Phase 1
- Bias in the Data Collection Phase 3
- Bias in the Analysis Phase 5
- Conclusion 6

Chapter 2: A Closer Look at Response Bias 8
- Sources 8
- Prevention 10
- Detection and Quantification 12
  - Data Collection Phase 12
  - Analysis Phase 13
- Adjustments 14
- Conclusion 17

Chapter 3: Simulation 20
- Introduction 20
- Methods 24
  - Sample Size 24
  - Model Parameters 24
    - Treatment Effect 25
    - Covariate 26
  - Reliability Distribution 26
  - Unreliable Response Style 27
  - Bias Correction Strategies and Models 28
  - Model Comparison 30
    - Organizational Structure 30
    - *Bias* and *RRMSE* (%) 31
    - Selecting the Best Strategy 32

Results 34
Prediction Equations 34
Subject- and Sample-Level Reliability 35
Comparison of Bias Correction Strategies 37
Recommendations 45
Comparison of Unadjusted and Adjusted Estimates 46
Conclusion 51

Chapter 4: Discussion 53
Limitations 56
Directions for Future Research 58

References 61

Appendices 69
Appendix A: R Code for Data Generation 70
Appendix B: R Code for Bias Correction Models 72
Appendix C: R Code for *Bias* and *RRMSE* (%) 89
Appendix D: R Code for *RRMSE* (%) Figures 90

List of Tables

Table 1: Summary of Response Bias Prevention, Detection/Quantification, and Adjustment Methods 18

Table 2: *RRMSE* (%) Margins by Scenario 33

Table 3: Prediction Equations, Log Odds of Outcome, and Probability of Outcome 34

Table 4: Subject-Level Reliability for 50% and 80% Reliable Samples 36

Table 5: *Bias* and *RRMSE* (%) for Competing Bias Correction Strategies, Opposite Responding 41

Table 6: *Bias* and *RRMSE* (%) for Competing Bias Correction Strategies, Arbitrary Responding 42

Table 7: *Bias* and *RRMSE* (%) for Competing Bias Correction Strategies, Socially Desirable Responding 43

Table 8: *Bias* and *RRMSE* (%) for Adjusted Social Desirability 44

Table 9: Recommended Bias Correction Strategies 45

Table 10: *Bias* and *RRMSE* (%) for Unadjusted and Adjusted Models, Opposite Responding 50

Table 11: *Bias* and *RRMSE* (%) for Unadjusted and Adjusted Models, Arbitrary Responding 50

Table 12: *Bias* and *RRMSE* (%) for Unadjusted and Adjusted Models, Socially Desirable Responding 51

Table 13: Step-by-Step Guide to Applying Bias Correction Strategies 55

List of Figures

Figure 1: Organizational Structure of Bias Correction Models 31

Figure 2: Reliability Distributions for 50% and 80% Reliable Samples (*N* = 20000) 35

Figure 3: Comparison of *RRMSE* (%) for Competing Bias Correction Strategies, Opposite Responding 38

Figure 4: Comparison of *RRMSE* (%) for Competing Bias Correction Strategies, Arbitrary Responding 38

Figure 5: Comparison of *RRMSE* (%) for Competing Bias Correction Strategies, Socially Desirable Responding 39

Figure 6: Comparison of *RRMSE* (%) for Unadjusted and Adjusted Models, Opposite Responding 47

Figure 7: Comparison of *RRMSE* (%) for Unadjusted and Adjusted Models, Arbitrary Responding 47

Figure 8: Comparison of *RRMSE* (%) for Unadjusted and Adjusted Models, Socially Desirable Responding 48

Abstract

Background: Response bias can distort treatment effect estimates and inferences in clinical trials. Although prevention, quantification, and adjustments have been developed, current methods are not applicable when subject-level reliability is used as the measure of response bias. Thus, the objective of the current study is to develop, test, and recommend a series of bias correction strategies for use in these cases. **Methods:** Monte Carlo simulation and logistic regression modeling were used to develop the strategies, examining the collective impact of sample size (*N*), effect size (*ES*), reliability distribution, and response style on estimating the treatment effect size in a series of hypothetical clinical trials. The strategies included a linear (LW), quadratic (QW), or cubic weight (CW) applied to the subject-level reliability; a reliability threshold (%); or a combination of the two (W-%). *Bias* and *percent relative root mean square error* (*RRMSE* (%)) were calculated for each treatment effect estimate and *RRMSE* (%) was compared to inform the bias correction recommendations. **Results:** The following recommendations are made for each *N* and *ES* combination: *N*=200/*ES*=small: no adjustment, *N*=200/*ES*=medium: 40%-LW, *N*=200/*ES*=large: 40%-QW, *N*=2000/*ES*=small: 40%-LW, *N*=2000/*ES*=medium: 55%-CW, *N*=2000/*ES*=large: 75%-CW, *N*=20000/*ES*=small: 70%-CW, *N*=20000/*ES*=medium: 85%-CW, *N*=20000/*ES*=large: 95%-CW. **Conclusion:** Employing these bias correction strategies in clinical trials where subject-level reliability can be calculated will decrease error and increase accuracy of estimates and validity of inferences.

Chapter 1: Overview of Bias

Bias is a systematic error that results in a sample statistic over- or underestimating a population parameter (Wackerly, Mendenhall, & Scheaffer, 2008), potentially leading to distorted results and inaccurate inferences. Statistically, bias is defined as:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta \qquad (1)$$

where $\hat{\theta}$ is the point estimate (i.e., sample statistic estimating the population parameter), $B(\hat{\theta})$ is the bias of the point estimate, $E(\hat{\theta})$ is the expected value (i.e., mean) of the point estimate, and θ is the population parameter; with positive $B(\hat{\theta})$ indicating overestimation and negative $B(\hat{\theta})$ indicating underestimation (Wackerly et al., 2008).

Several types of bias exist and can be introduced by a myriad of practices throughout the research process, typically in the design phase, during data collection, and/or during the analysis phase. In the sections that follow, various biases will be reviewed within each of these study phases.

Bias in the Design Phase

The primary bias introduced during the design phase of a study is *selection bias*, wherein the study sample differs from the population that they are intended to represent, potentially leading to results that struggle with generalization (L. K. Alexander, B. Lopes, K. Ricchetti-Masterson, & K. B. Yeatts, 2015a; Berk, 1983; Hultsch, MacDonald, Hunter, Maitland, & Dixon, 2002; Pruchno et al., 2008). More specifically, selection bias occurs when certain groups have a higher probability of being chosen to participate in the study than others (Berk, 1983). Convenience samples are particularly prone to this bias, as subjects are chosen from populations that researchers have easy access to such as students at a particular university or patients from local hospitals in a single metropolitan area. Indeed, these subjects often share similar traits as a consequence of being clustered within a shared environment (Hultsch et al., 2002; Pruchno et al., 2008). Case-control studies are also vulnerable to selection bias wherein different procedures are used to select cases than are used to select controls, resulting in differing selection probabilities for each group (L. K. Alexander et al., 2015a).

Self-selection bias is a special case of selection bias, wherein some individuals are more likely to participate in research studies compared to others, potentially resulting in differing underlying characteristics between those who participate and those who do not. This phenomenon is also known as *volunteer bias* (L. K. Alexander et al., 2015a; Heckman, 1979; Krishna, Maithreyi, & Surpaneni, 2010). Self-selection bias also occurs when rates of attrition differ among sample groups; that is, some subjects are more likely to dropout of a study compared to others (L. K. Alexander et al., 2015a).

Non-respondent bias is another subtype of selection bias, which acutally occurs during data collection, where subjects who fail to respond to a question or to a survey significantly differ from subjects who do respond to the question/survey (Krishna et al., 2010). For example, research has shown that questions regarding subject income are vulnerable to item non-response, such that subjects at each end of the socioeconomic status distribution may be less likely to report their income compared to subjects in the middle (Juster & Kuester, 1991; Pfeffer & Griffin, 2017; Riphahn & Serfling, 2005; Turrell, 2000).

Bias in the Data Collection Phase

Several biases occur during the data collection phase, with most falling under the category of *information bias*, or *measurement bias*, which arises when data are measured or classified inaccurately (L. K. Alexander, Brettania Lopes, Kristen Ricchetti-Masterson, & Karin B. Yeatts, 2015b). Some specific subtypes of information bias are misclassification, interviewer, contamination, co-intervention, compliance, and response biases (L. K. Alexander et al., 2015b; Cox et al., 2009; Delgado-Rodriguez & Llorca, 2004; Krishna et al., 2010; Pannucci & Wilkins, 2010; Sackett, 2007; Tripepi, Jager, Dekker, Wanner, & Zoccali, 2008).

Misclassification bias is the inaccurate collection and utilization of classification data that results in subjects in a non-randomized study being incorrectly assigned to a specific subgroup (e.g., exposed or non-exposed; diseased or non-diseased). Furthermore, misclassification bias can be either differential or non-differential. Differential misclassification occurs when subjects in one subgroup are misclassified more frequently than those in another, but this occurs equally among subgroups under non-differential misclassification (L. K. Alexander et al., 2015b; Cox et al., 2009; Tripepi et al., 2008).

Interviewer bias occurs when the researcher collecting the data subconsciously influences the subject's responses (Tripepi et al., 2008) or when differences in obtaining, recording, or interpreting information between groups occurs (L. K. Alexander et al., 2015b; Pannucci & Wilkins, 2010). Blinding, or withholding group status from the interviewer is the best approach to avoid this type of bias (Tripepi et al., 2008).

Contamination bias results from the control group unintentionally receiving the experimental intervention, thus contaminating their control status and potentially decreasing the difference in outcomes for the experimental and control groups (Krishna et al., 2010). Although cluster randomization – wherein groups of subjects are randomized to a treatment arm, as opposed to individual randomization – has been suggested to minimize contamination bias, many researchers have reported significant flaws with this method (Giraudeau & Ravaud, 2009; Hahn, Puffer, Torgerson, & Watson, 2005; Keogh-Brown et al., 2007; Torgerson, 2001).

Cointervention bias occurs when a study participant is receiving additional care outside of the study that other subjects are not receiving, as this can impact the study intervention. This often occurs for control subjects who have other clinicians or caregivers providing additional care, knowing that the subject is a control in the study, to ensure that their condition improves (Krishna et al., 2010; Sackett, 2007, 2011). Blinding subjects to treatment condition is one way to prevent cointervention bias (Schulz & Grimes, 2002).

Compliance bias is when some subjects in an intervention study adhere to the treatment regimen more strictly than others, potentially distorting estimates of intervention efficacy (Delgado-Rodriguez & Llorca, 2004; Krishna et al., 2010). This is typically evidenced when subjects drop out of a study before they complete the entire intervention or when subjects fail to complete portions of the intervention that they are instructed to complete. Clearly, non-compliance can result in substantial missing data, which can have negative impacts if not handled properly, including decreased statistical power (Melnikow & Kiefe, 1994), inaccurate conclusions on drug dosage or safety (Little et al., 2012), or other inaccurate inferences of treatment comparisons (Myers, 2000).

Response bias, otherwise known as *respondent bias* or *self-reporting bias*, is a broad area of biases wherein the subject provides unreliable responses to questions, which may be due to a variety of factors, both conscious and subconscious (Althubaiti, 2016; Lavrakas, 2008b; Szklo & Nieto, 2014). Indeed, response bias consists of several subtypes of biases, including social desirability, demand characteristics, extreme responding, acquiescence, careless/random responding, and recall bias (Althubaiti, 2016; Fadnes, Taube, & Tylleskar, 2009), all of which will be discussed and elaborated on in Chapter 2.

Bias in the Analysis Phase

During the analysis phase of a study, the primary bias to emerge is *confounding*, where a certain variable is associated with both the exposure (i.e., predictor) and the outcome, and so appears to be influencing the exposure-outcome relationship, but does not actually serve a relational purpose (Cox et al., 2009; Grimes & Schulz, 2002; Pannucci & Wilkins, 2010). Indeed, confounding distorts the observed relationship between an exposure and an outcome, such that an association is detected where one does not truly exist, an association fails to be detected when one does truly exist, or the association appears to be weaker or stronger than truly exists (Braga, Farrokhyar, & Bhandari, 2012). A confounding variable has three defining characteristics: (1) predictive of outcome but not caused by exposure or outcome variables (i.e., serves as a risk/predictive factor for outcome); (2) associated with exposure, such that rates of the confounder differ among exposed and unexposed groups; and (3) not an intermediate step on the causal pathway between exposure and outcome (Skelly, Dettori, & Brodt, 2012). The effects of confounding are alleviated by randomization since the covariate will, theoretically, be equally distributed among each group, given a large enough sample size, and thus primarily affects observational designs (Braga et al., 2012; Pourhoseingholi, Baghestani, & Vahedi, 2012). However, methods are available for observational studies that aim to reduce the impact of confounding, such as the use of propensity scores (D'Agostino, 1998); setting inclusion criteria that limit the range of the confounding variable (e.g., certain age group or specific gender); using stratified analyses, which allow results to be compared for the subgroups corresponding to the various levels of the confounder; and by using multivariate models, wherein the confounding variable is controlled for during statistical modeling (Skelly et al., 2012).

Publication bias also appears at the end of a study and occurs when significant results are published, and insignificant results fail to be published (Joober, Schmitz, Annable, & Boksa, 2012; Tripepi et al., 2008). This type of bias occurs on the part of both researchers and journal editors. For researchers, obtaining grant funding and academic career promotions are both highly competitive and often depend upon publishing studies with significant results. For journal editors, obtaining a high citation index is also highly competitive and depends on publishing significant results, since studies with insignificant results are less likely to be cited by researchers than significant ones (Joober et al., 2012).

Conclusion

Although this chapter reviews the primary biases encountered in clinical and epidemiological research, it by no means serves as a comprehensive review. For a more exhaustive review of bias in research, please refer to Delgado-Rodriguez and Llorca (2004), Krishna et al. (2010), and Tripepi et al. (2008).

Each aforementioned bias poses a risk to statistical analyses and inferences, individually or in a compounded fashion. That is, some biases may cause or contribute to one or more other biases, potentially exponentiating the distorted estimates. Response bias, for instance, can contribute to misclassification bias, which can, in turn, lead to the emergence of confounding.

Indeed, each type of bias, or combination of biases, has the potential to skew the estimated value either closer to or further from the true value. That is, estimates of associations, treatment effects, drug doses, etc. can either under- or overestimate their true values, resulting in distorted estimates that can lead to inaccurate interpretations and inferences. These misleading conclusions have the potential to cause adverse events, especially in the case of drug safety if a dosing regimen is set too low (i.e., drug may be ineffective) or too high (i.e., toxicity), or if side effects are not accurately reported. Furthermore, resources, namely grant funding, may not be distributed in a way that produces the best outcomes if the true effects are not evident due to biased results and/or reporting. Thus, it is vital to minimize bias by selecting the appropriate study design; implementing randomization and blinding when possible; and utilizing proper statistical techniques to adjust for missing data, confounders, and other bias-inducing factors during the analysis phase.

Chapter 2: A Closer Look at Response Bias

As briefly discussed in Chapter 1, *response bias*, otherwise known as *respondent bias* or *self-reporting bias*, is a broad area of biases wherein a subject provides unreliable, inaccurate, or dishonest responses to questions, resulting in statistical estimates either over- or underestimating population parameters (Althubaiti, 2016; Lavrakas, 2008b; Szklo & Nieto, 2014; Wackerly et al., 2008).

Sources

The specific biases included under the umbrella term of response bias are social desirability, demand characteristics, extreme responding, acquiescence, careless/random responding, and recall bias (Althubaiti, 2016; Fadnes et al., 2009; Meyer, Faust, Faust, Baker, & Cook, 2013).

Social desirability is one of the most common types of response bias (Nederhof, 1985) and is typically observed in the context of attitudes, beliefs, and behaviors regarding sensitive issues such as sex, drugs, crime, racism, obesity, and many others (Elgar, Roberts, Tudor-Smith, & Moore, 2005; Krumpal, 2013). Indeed, social desirability occurs when subjects respond to a question according to social norms, which typically paints the subjects in a favorable light (van de Mortel, 2008). That is, subjects often overreport socially accepted beliefs/behaviors and underreport socially undesirable ones (Krumpal, 2013).

Similarly, *demand characteristics*, or the good-subject effect, arise when subjects respond/behave in a manner that they perceive the researcher to expect from them. That is, the characteristics of the researcher demand a specific response/behavior from the subject (Nichols & Maner, 2008; Orne, 1962).

Extreme responding has two forms, categorial and continuous. In the categorical form, this type of bias is characterized by a subject using only the end response options on a rating scale, failing to utilize the middle options (Batchelor & Miao, 2016). In the continuous form, this bias appears as values that are highly improbable or impossible. *Acquiescence* is a special case of extreme responding that occurs when a subject provides positive/affirmative answers to questions, regardless of their content (Hinz, Michalski, Schwarz, & Herzberg, 2007; Lavrakas, 2008a).

Careless responding occurs when subjects provide random answers to questions, regardless of directionality (affirmative, contradictory, or neutral), without considering the question content (Meyer et al., 2013).

Recall bias is one that may occur subconsciously and results from the subject's inability to accurately recall a past event. This type of bias is more frequently observed in epidemiologic studies, particularly within case-control or retrospective cohort designs (Althubaiti, 2016), which require respondents to recall prior exposure histories that may have occurred several years or even decades in the past. Indeed, recall bias is known to increase as time between assessment and recalled event increase (Fadnes et al., 2009).

In addition to these specific types of biases, response bias can also result from question wording and response format. For instance, Brener, Grunbaum, Kann, McManus, and Ross

(2004) conducted a study on question wording and found that nearly 40% of question sets – where two questions ask the same thing, but one is phrased differently from the other – experienced significantly different prevalence estimates. Additionally, Cabooter, Weijters, Geuens, and Vermeir (2016) evaluated the effect of scale formats on subject responses and found that subjects interpret scales differently depending on whether labels are used at one or both ends (polarity) and whether only positive numbers are used, or both positive and negative numbers. Cross-cultural differences can also contribute to response bias (e.g., T. Johnson, Kulesa, Cho, & Shavitt, 2005; Lai, Cummins, & Lau, 2013) since questions are not necessarily interpreted or responded to the same way across cultures (Iwata, 2014).

Unfortunately, unreliable responses within self-reports are common and the impacts on statistical estimates are well documented (Adams, Soumerai, Lomas, & Ross-Degnan, 1999; Hebert et al., 2002; Krumpal, 2013; Mazor, Clauser, Field, Yood, & Gurwitz, 2002; Meyer et al., 2013; Navarro-González, Lorenzo-Seva, & Vigil-Colet, 2016; Preston, Fishman, & Stokes, 2015; Shields, Gorber, & Tremblay, 2008; van de Mortel, 2008).

Prevention

Many methods have been utilized to prevent response bias from occurring, or at least to reduce its impact on statistical estimates. For instance, Nederhof (1985) recommends using neutral questions and Wouters, Maesschalck, Peeters, and Roosen (2014) recommend using anchors at each end of a categorical response spectrum rather than at each response option. Although this latter option seems counterintuitive, studies have shown that subjects may interpret the labels differently, thus introducing bias (Wouters et al., 2014).

Another approach to preventing response bias is the use of survey administration modes that foster subject anonymity, as research has shown these modalities to result in decreased bias as compared to others. For example, one study administered psychological surveys face-to-face, over the phone, or on a computer (online and offline) and found that responses differed significantly between computer and face-to-face surveys, computer and phone surveys, and face-to-face and phone surveys (Zhang, Kuchinke, Woud, Velten, & Margraf, 2017). Another study found that surveys completed by mail had a wider range of responses compared to surveys completed via phone (Hall, 1995). Conversely, a meta-analysis conducted on paper versus online surveys found no significant differences in responses across survey modalities (Dodou & Winter, 2014), but this may be because both modes (paper and online) allow privacy/anonymity while completing the survey. Indeed, a study by Ong and Weiss (2000) showed that prevalence estimates of sensitive behaviors were significantly greater when anonymity was provided compared to offering only confidentiality, and a review by Tourangeau and Yan (2007) discussed the advantage of self-administered surveys in a private setting producing less biased responses.

The bogus pipeline is also used to prevent response bias, where a fake polygraph-type device is connected to the subject and he/she is told that it will detect dishonest responses, even though it is not actually collecting any data. In this context, subjects must choose between offering a potentially socially undesirable response or looking like a liar (also socially undesirable; Aguinis, Pierce, & Quigley, 1993; Nederhof, 1985; Roese & Jamieson, 1993).

The randomized response technique (RRT) is another approach that is used in attempts to reduce or eliminate response bias due to social desirability (Greenberg, Kuebler, Abernathy,

& Horvitz, 1971; Hsieh, Lee, Li, & Tu, 2016; Nederhof, 1985; Warner, 1965). In this approach, the subject uses a randomization device, such as a coin toss, to determine which of two questions to answer (one about the sensitive topic, the other being non-sensitive), with the interviewer being blinded to the question being answered. See Warner (1965) and Greenberg et al. (1971) for details.

Regarding recall bias, using short recall periods and recalling events that occur infrequently are likely to result in decreased bias as opposed to longer recall periods and recall of more frequent events. Indeed, research has shown that recall accuracy deteriorates as the recall period increases and subjects have difficulty pinpointing exactly when a specific instance of an event occurred if the event itself occurred several times (Althubaiti, 2016).

Detection and Quantification

Even with a variety of methods available to prevent/reduce response bias, they are not always feasible to implement, and even if they are, no approach is perfect, so it is vital to be able to detect and quantify response bias. Several methods have been developed for this purpose that are implemented either during the data collection phase or the analysis phase of the study.

Data Collection Phase

During the data collection phase, certain types of response bias can be measured by using a social desirability detection scale (Lambert, Arbuckle, & Holden, 2016; O'Leary, Diller, & Recklitis, 2007), with two of the most common being the Balanced Inventory of Desirability Responding (Lanyon & Carle, 2007; Paulhus, 1998) and the Marlowe-Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960; Loo & Loewen, 2004).

Another common approach is to ask the subject to rate his/her level of honesty, either overall, or for a specific question or survey. For example, some researchers have asked subjects "Overall, how honest would you say you were in answering this questionnaire?" with a 5- to 7-point Likert scale ranging from 1 = completely honest to 5/7 = not honest at all (Siegel, Aten, & Roghmann, 1998; Wiederman, 1997; Zimmerman & Langer, 1995).

Similarly, survey questions can be designed to detect careless/random responding (e.g., "I read this item before answering") and inserted into a given questionnaire. Placing these items intermittently throughout the survey, rather than all at the end, allows for more accurate detection of when the careless responding began (Meyer et al., 2013).

Analysis Phase

Perhaps the most commonly used approach to detect and quantify response bias during the analysis phase is identifying inconsistent reporting across question sets (Zimmerman & Langer, 1995), time points (Langeland et al., 2015; Toneatto, Sobell, & Sobell, 1992; Zimmerman & Langer, 1995), or data collection methods (Elgar et al., 2005; Griesler, Kandel, Schaffran, Hu, & Davies, 2008; Hebert et al., 2002; Shields et al., 2008). For example, significant differences between self-reported weight and objectively measured weight would indicate inconsistent reporting across data collection methods. Identifying response bias in this manner is becoming common practice in the height/weight/BMI research area since it is a fairly simple process to measure height and weight to calculate BMI and to also collect this information via self-report surveys (e.g., Brault, Turcotte, Aimé, Côté, & Bégin, 2015; Clarke, Sastry, Duffy, & Ailshire, 2014; Gorber, Shields, Tremblay, & McDowell, 2008; Ward et al., 2016).

Another approach is to identify excessive or extreme reporting. For example, research has shown that men tend to report greater numbers of sexual experiences/partners compared to women (e.g., M. G. Alexander & Fisher, 2003; Beaussart & Kaufman, 2013; Smith, 1991; Wiederman, 1997). In the context of lifetime number of sex partners, Wiederman (1997) quantifies extreme reporting as values that fall outside of the statistically normal range (i.e., 10 partners).

Lastly, Hinz et al. (2007) identified and quantified response bias in their study by using an outcome measure that was equally balanced with positively and negatively phrased questions to create an acquiescence score. The scale was summed, without inverting negatively phrased items, and had a range from 0 to 100 with 50 as the median. Individuals scoring at either extreme evidenced strong acquiescence either positively or negatively.

Adjustments

Once response bias has been identified and quantified, adjustments can be made to reduce or eliminate the bias, resulting in more accurate statistical estimates. One approach is to simply identify predictors of biased responding and control for those variables during the modeling process. This approach has led to mixed results as its effectiveness depends on a sufficient number of bias predictors being adequately measured. That is, if too few bias predictors are identified and controlled for, or if the predictors are not measured accurately, the bias may only be reduced by a negligible amount, if at all, resulting in none to minimal improvements in estimator accuracy.

Gorber et al. (2008) experienced significant bias reduction using this approach to adjust for differences in directly measured and self-reported height and weight in the 2005 Canadian

Community Health Survey sample. First, they calculated bias by subtracting the self-reported height/weight from the directly measured height/weight. Then, they determined the factors predictive of bias by using multiple regression models with demographic and health variables as predictors. Next, bias correction models were constructed by using the measured value of height/weight as the outcome with the self-reported value of height/weight and the significant predictive factors determined in the last step as the predictors. Indeed, the adjusted estimates using the bias correction models experienced significant improvement, as they were not statistically different from the directly measured values. See Gorber et al. (2008) for details.

On the other hand, Lauritsen and Swicegood (1997), who investigated age at first sexual intercourse in a national sample of adolescents, did not experience improved estimator accuracy with this method. The researchers identified and controlled for the following variables as predictors of inconsistent reporting: age, gender, race, grade point average, family structure, household income, and neighborhood condition. Although several bias predictors were controlled for in their model, the estimate did not change significantly, indicating that this approach to bias-adjustment was ineffective in their study.

Another approach to response bias correction was developed by Zimmerman and Langer (1995), who examined sexual behaviors in a sample of tenth grade students. The researchers collected information on subject-level reliability from each subject using the self-reported honesty 5-point Likert scale discussed previously. Subjects were given a weight of 0.0 if they claimed to be “not honest at all” or “not completely honest” and were dropped from the analyses. Subjects who reported that they were “fairly honest” received a weight of 0.33, “very honest” a weight of 0.67, and “completely honest” a weight of 1.0, indicating that their responses had a minimal, moderate, and large influence on generating the estimate, respectively. However, the implementation of this weighting scheme did not result in a significant change in the estimates.

Another response bias adjustment involves the use of an Expectation-Maximization-Bayesian (EMB) algorithm, which was developed by researchers estimating the prevalence of rape and domestic violence using the National Crime Victimization Survey (Yu, Stasny, & Li, 2008). The EMB model incorporates prior non-time-sensitive information into a model that accounts for factors that contribute to biased self-reports. In this study, the EMB model included type of crime, whether the spouse was present during the data collection interview, and whether the interview was over the phone or in person (researchers assume that presence of spouse and interview via phone increases response bias). Use of the Bayesian model significantly improved estimates compared to the frequentist model. See Yu et al. (2008) for details.

The exponential decay model (EDM), which was used by researchers examining initial age of alcohol and marijuana use by adolescents in a nationwide sample, is another model that has been used for response bias correction (R. A. Johnson, Gerstein, & Rasinski, 1998). In their study, the researchers found that as time increased between first use and data collection period, estimates of alcohol/marijuana incidence decreased. Thus, an EDM was implemented to adjust for this response bias, statistically increasing accuracy of incidence estimates.

Another study used a non-parametric statistical matching algorithm to adjust self-reported height/weight in the Behavioral Risk Factor Surveillance System (BRFSS) using the relationship between self-reported and measured height/weight in the National Health and

Nutrition Examination Survey (NHANES; Ward et al., 2016). Subjects in each study were matched on height and weight percentiles within various demographic subgroups (age, gender, race, etc.) and the statistical matching algorithm was applied. This method resulted in significantly improved adjusted BRFSS height/weight estimates, as they did not significantly differ from the directly measured NHANES height/weight values.

Score standardization has also been implemented to address bias from extreme responding (Brinker, 2002) and cross-cultural response bias (Fischer, 2004), resulting in more accurate estimates. On the subject-level, this method involves (1) transforming a subject's raw scores across variables to z-scores by standardizing the distribution of their responses (Brinker, 2002) or (2) using deviation scores, which are derived via ipsatization (Fischer, 2004). On the group level, this method involves group mean centering, or subtracting the group mean from the subject's score, and on the cultural level, grand mean centering (Fischer, 2004).

Conclusion

Response bias is a significant problem in survey research and the impacts on statistical estimates can be highly influential, potentially leading to distorted results and inaccurate inferences. Understanding the various types of response bias and how they can be minimized is critical to obtaining and maintaining valid results in research studies. The objective of this chapter is to collect and synthesize information on the prevention, detection, quantification, and adjustment methods for various types of response bias in order to assist researchers utilizing self-report surveys in increasing the validity of their findings. Table 1 presents a summary of these various methods and each type of response bias to which they apply.

Table 1: Summary of Response Bias Prevention, Detection/Quantification, and Adjustment Methods

Method Type	Method	Applicable Response Bias Type
Prevention	Neutral questions	Any
	End anchors	Any
	Settings that permit anonymity	Social desirability
	Bogus Pipeline	Social desirability
	Randomized Response Technique (RRT)	Social desirability
	Short recall periods	Recall bias
	Recall infrequent events	Recall bias
Detection / Quantification	Evaluate predictive factors for inconsistent reporting	Any
	Self-reported honesty items	Any
	Compare objective and subjective measures	Any
	Social desirability scales	Social desirability
	Item detection	Careless responding
	Extreme values outside statistically normal range	Extreme responding
	Do not invert negatively phrased items	Extreme responding Acquiescence
Adjustment	Control for factors predictive of biased responding	Any
	Weights based on self-reported honesty	Any
	Statistical matching	Any
	Standardization	Any
	Expectation-Maximization-Bayesian (EMB) algorithm	Social desirability
	Exponential Decay Model (EDM)	Recall bias

Before designing a study, it is vital for researchers to assess the various types of bias that their study is vulnerable to, and methods to minimize, detect/quantify, and adjust for each of them individually. For example, case-control studies are particularly vulnerable to selection bias, observer bias, and misclassification bias (among others) and researchers should take every step they can to implement various methods to prevent and minimize the impacts of these potential biases in order to increase the validity of their findings.

Implementing bias minimization efforts requires early planning and utilization of various methods through each stage of the research process (i.e., study design, data collection, data analysis, and publication). Although extra time and effort are required to effectively and efficiently implement these methods, generating minimally biased results will provide more accurate findings, increasing the validity and integrity of the research produced. Indeed, investigators should exert the effort required to produce the most valid results, enhancing the quality of research disseminated to the scientific community.

Chapter 3: Simulation

Introduction

Self-report surveys are one of the most common methods of collecting data in research studies (Saczynski, McManus, & Goldberg, 2013) due to the relative ease of collecting information and the ability to assess experiences, attitudes, and beliefs that cannot be obtained from other data collection formats. However, self-reported data has a known vulnerability to response bias, which can lead to distorted estimates and misleading inferences (Adams et al., 1999; Mazor et al., 2002; van de Mortel, 2008). Indeed, response bias, otherwise known as respondent bias or self-reporting bias, is a broad area of biases wherein a subject provides unreliable, inaccurate, or dishonest responses to questions, resulting in statistical estimates either over- or underestimating population parameters (Althubaiti, 2016; Lavrakas, 2008b; Szklo & Nieto, 2014; Wackerly et al., 2008).

Perhaps the most commonly used approach to detecting response bias in self-report survey research is identifying inconsistent reporting across question sets (Zimmerman & Langer, 1995), time points (Langeland et al., 2015; Toneatto et al., 1992; Zimmerman & Langer, 1995), or data collection methods (Elgar et al., 2005; Griesler et al., 2008; Hebert et al., 2002; Shields et al., 2008). For example, if a question asks, “Have you ever had sex before?” and the subject says “No,” but then answers “Yes” to another question that asks, “Have you ever been pregnant before?” this would indicate inconsistent reporting across question sets. If an item asks, “Have you ever had sex before?” and the subject says “Yes” at baseline, but “No” at follow-up, this would constitute inconsistent reporting across time points. Or if a question asks, “Have you ever had sex before?” and the subject says “No,” but then a blood or urine test reveals that she is pregnant, this would be an example of response bias across data collection methods.

Identifying response bias through inconsistent self-reports allows researchers to estimate subject- and sample-level reliability. Subject-level reliability can be estimated by calculating the proportion of inconsistent responses observed for each individual and subtracting this value from 1. For example, if there are 10 sets of questions where inconsistent reporting is possible, and a subject provides inconsistent reports in 2 of these question sets, then the proportion of inconsistent responses would be $\frac{2}{10}$ = 0.20 or 20% and the subject-level reliability would be 1 – 0.20 = 0.80 or 80%. This can also be thought of as the subject’s probability of responding honestly. Sample reliability can then be calculated in one of two ways. The first option would be to compute the average subject-level reliability across the entire sample and the second option would be to calculate the proportion of reliable responders (i.e., subjects with 0 illogically inconsistent observations). For instance, if 25 out of 100 (0.25) subjects in a sample had at least one inconsistent response, then the sample-level reliability would be 1 – 0.25 = 0.75 or 75%. Response bias that is identifiable and quantifiable using the latter approach is part of the foundation for the current study.

The impact of inconsistent reporting is evidenced in two school-based intervention studies aimed at reducing risky sexual behaviors in adolescents (Walsh-Buhi et al., 2016; Zimmerman & Langer, 1995). The first is the Teen Outreach Program (TOP), which was conducted on two cohorts of high school freshman across 28 public schools throughout the state of Florida from 2012 – 2014, with a total sample size of 7,976. At each of three time points, subjects were asked if they had ever had sex, and if they had ever been pregnant (females) or if they had ever gotten someone else pregnant (males). For cohort one, 3,621 subjects responded, with 105 subjects answering “Yes” at baseline to having ever had sex, but at a later time point answering “No.” Thus, 2.90% of subjects provided inconsistent responses to this item. Similarly, 19 subjects answered “Yes” to having ever been pregnant or gotten someone else pregnant but later answered “No,” resulting in 0.52% of the sample providing inconsistent responses to this item (Walsh-Buhi et al., 2016). A similar phenomenon was seen in the AIDS Education Program, which was conducted on 1,886 high school sophomores across 8 schools in Miami-Dade, Florida during the early 1990s. During at least two of the four time points, subjects were asked if they had ever engaged in a variety of sexual and substance use-related behaviors. The highest rate of inconsistent reporting in this study was regarding having ever engaged in sexual intercourse with a same-gender partner. Of those who said “Yes” to engaging in this behavior, nearly 50% later said they had never done so. The high level of inconsistent reporting in this example highlights the severe bias that can occur in the context of sensitive topics (i.e., those vulnerable to high levels of social desirability). Other items that evidenced illogically inconsistent reporting in this study: ever had sexual intercourse (approximately 10% of those who initially said “Yes,” later said “No”), ever used marijuana (14%), ever drank alcohol (8%), and ever used cigarettes (8%; Zimmerman & Langer, 1995). Please note that the rates of inconsistent reporting were calculated differently for each of these two studies. The former study calculated inconsistent reporters from the *entire* sample (including those who said yes to the item at baseline and those who said no), whereas the latter study calculated inconsistent reporting from only subjects who said yes to the item at baseline (excluding subjects who said no).

Reviewing the literature for relevant bias correction strategies reveals a gap that quickly needs to be filled. Indeed, many bias correction methods have been developed, as discussed in Chapter 2, but none apply to the specific situation of having inconsistent reporting as the method for measuring response bias or when historical data is present that provides information on subject-level reliability. Even for the latter school-based intervention example above, the authors, Zimmerman and Langer (1995), used self-reported honesty, not inconsistent reporting, to assign weights to subjects' responses such that greater weights were given to respondents with higher levels of reported honesty. Indeed, their purpose for evaluating inconsistent reporting within their sample was to validate their self-reported honesty data.

Consequently, the current study proposes adjustment methods that are applicable in situations where subject-level reliability is calculated through inconsistent reporting. Specifically, bias correction strategies were developed and tested using Monte Carlo simulation and logistic regression modeling. The models are a function of sample size, effect size, reliability distribution, and unreliable response style and, in the current study, were used to estimate the treatment effect for a series of hypothetical clinical trials. *Bias* and *percent relative root mean square error* (*RRMSE* (%)) were calculated for each treatment effect estimate and *RRMSE* (%) was compared among the various models to inform the selection of best strategies for a variety of study scenarios. Recommended strategies are proposed for clinical trials with various combinations of sample and effect sizes, and the application of these strategies to real-world datasets are discussed.

Methods

Simulation was used to develop the bias correction strategies, which allows for the generation of "true" parameters in order to examine the accuracy of model estimates. Four factors that vary across research settings were investigated at various levels to assess their collective impact on the treatment effect estimates, including sample size, effect size, reliability distribution, and unreliable response style, each of which are discussed below.

Sample Size

Three samples were simulated of sizes *N* = 200, 2000, and 20000, each with half of the sample assigned to the treatment group and the other half, the control group. These sample sizes were chosen to reflect common real-world sample sizes and to evaluate how the bias correction models behave within a variety of sample sizes, which will help inform the corresponding bias correction recommendations.

Model Parameters

Bias correction models were developed using logistic regression modeling, with a binary outcome (0 = no, 1 = yes), a binary treatment condition (0 = control, 1 = treatment), and a binary covariate (0 = group A, 1 = group B). The simulated model is as follows:

$$logit\ [\ P\ (Y = 1)\ ] = \beta_0 + \beta_1\ (treatment) + \beta_2\ (covariate)$$

where Y is the outcome of interest, β_0 is the log odds of the outcome for the reference group (i.e., intercept), β_1 is the difference in the log odds of the outcome between the treatment and control groups (i.e., treatment effect size) when the covariate is held constant, and β_2 is the difference in log odds of the outcome between groups A and B when treatment is held constant. The “true” intercept was simulated at -1.0, which corresponds to the outcome occurring in approximately 27% of the reference group (control, group A). The simulation of the “true” treatment effect and covariate parameters are discussed below. Once these parameters were simulated, they were linearly combined and set equal to the outcome variable, which was then run through the inverse logit function, resulting in the probability of outcome (p). This probability was then used to simulate the “true” outcome via random generation from the binomial distribution:

$$p(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \qquad y = 0, 1 \quad and \quad 0 \leq p \leq 1 \qquad (2)$$

where y is a specific level of the outcome (0 or 1), n is the number of times the outcome was generated (1), and $p(y)$ is the probability of y.

Treatment Effect

The modeled treatment effect is an odds ratio (*OR*) that represents the difference in odds of saying “yes” to the outcome question for those in the treatment group versus those in the control group. Due to the logit nature of the model, log odds ratios were used. Specifically, log odds ratios of -0.3795 (*OR* = 0.6842), -0.9152 (*OR* = 0.4004), and -1.4204 (*OR* = 0.2416) were used in the simulation, with inverse *OR*s of 1.4615, 2.4972, and 4.1387. These inverse *OR*s correspond with Cohen’s *d* treatment effect sizes (*ES*) of 0.2 (small), 0.5 (medium), and 0.8 (large) when the outcome of interest is present in 10% of the unexposed group (in this case, the control group). See Chen, Cohen, and Chen (2010) for a detailed comparison of and conversion between *OR*s and Cohen’s *d* effect sizes.

Covariate

To mimic real-world differences in unreliable reporting, a binary self-reported covariate was generated wherein group A is more reliable than group B, which is controlled for in the model. This variable can be thought of as a predictor of bias and also as a risk factor that some people have at baseline and others do not (e.g., subjects who have drank alcohol before versus those who have not). A log odds ratio of 0.75 (*OR* = 2.12) was simulated for the covariate, suggesting that, when treatment is held constant, the odds of the outcome for subjects in the less reliable group (group B) are 2.12 times larger than the odds of the outcome for subjects in the more reliable group (group A).

Reliability Distribution

The reliability distribution consists of the proportion of the sample that provides accurate, honest responses (i.e., reliable) and the proportion that provides inaccurate responses (i.e., unreliable). In this study, the reliability proportions were simulated at approximately 50% and 80% (making unreliable proportions equal to 50% and 20%, respectively). Another approach to generating the sample-level reliability would have been to simply calculate the average subject-level reliability.

The first step in generating the reliability distributions was to randomly generate absolute values for each subject using the normal distribution:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} \qquad -\infty < y < \infty \tag{3}$$

where $f(y)$ is the absolute value to be generated, μ is the mean (*M*), and σ is the standard deviation (*SD*). Values *M* = 2 (for ~50% reliability) or *M* = 1 (for ~80% reliability) and *SD* = 1 were used. Then, subjects in group B had their reliability reduced by 25% by multiplying their absolute values by 0.75. The next step consisted of reducing absolute values > 1 down to 1 so that all values were contained within the boundaries of probability (0 to 1). Lastly, the absolute values were converted to a binary scale (0 = unreliable and 1 = reliable) for each subject using the binomial distribution (random generation) with probability of an accurate response (p) equal to the absolute value from the last step. That is, each subject's value (0 or 1) categorizes them as either a reliable or unreliable reporter based on their probability of providing accurate responses.

Unreliable Response Style

Three response styles were simulated to reflect the nature of unreliable responding observed in survey research. These response styles are implemented in cases where the subject was categorized as an unreliable reporter, as just discussed, to incorporate biased self-reports. The model has two variables that are vulnerable to response bias, outcome and covariate, and the same unreliable response style was implemented for both variables.

The first unreliable response style is social desirability, where the subject's response (self-report) reflects the social norm for the context in question. Examples of questions vulnerable to social desirability include "Do you exercise regularly?" or "Have you ever been a victim of rape or incest?" In this study, unreliable responses under this response style are coded as 0, making the outcome (y = 1) a negative belief/behavior. The amount of bias that this unreliable response style contributes is dependent upon the prevalence of true 0 responses in the data (or 1 if that were the more socially desirable response). If the majority of subjects have a true response of 0, then socially desirable responding contributes relatively little bias since the data do not change very much by incorporating the biased self-reports. However, if most true responses are 1, then more bias is contributed since more values will change from 1 to 0. In the current study, the prevalence of 0 in the data ranged from 56 – 73% in the control group, thus, socially desirable responding contributed relatively little bias overall.

The second unreliable response style is arbitrary responding, otherwise known as careless responding. This occurs when a subject chooses an answer randomly, disregarding the question content. In this study, arbitrary responses were randomly generated at 0 or 1, with each having a 50% likelihood of generation. Therefore, arbitrary responding contributed a moderate amount of bias to the data.

The final unreliable response style is opposite responding, where the subject chooses the answer that is opposite to the true response. For example, if the question is "Are you an only child?" and the true answer is Yes (code = 1), then the unreliable response is No (code = 0), and vice versa. Since responses were dichotomized, opposite response was coded as (1 – response) resulting in a 0 if the true response was 1, or a 1 if the true response was 0. Since opposite responding always provided an unreliable response, this response style contributed bias 100% of the time, and provided the greatest amount of bias of all three response styles.

Bias Correction Strategies and Models

Bias correction strategies were evaluated using three primary methods: reliability weights, reliability thresholds, and weight/threshold combinations. The weighting method provided larger weights for subjects with higher probabilities of honest responding, thus allowing reliable responses to have a larger influence on the treatment effect estimates than less-reliable responses. The subject's probability of responding honestly (p) was used as the weight such that weights ranged from 0 to 1. The specific weighting methods utilized consisted of linear weighting (p), quadratic weighting (p^2), and cubic weighting (p^3). If, for example, $p = 1$, then $p^2 = 1$, and $p^3 = 1$, but if $p = .5$, then $p^2 = 0.25$, and $p^3 = 0.125$, showing that reliable responses have a consistently large influence on generating the estimate, but that unreliable responses have their level of influence minimized as the weighting method intensifies. Comparing these two reliabilities, it is clear the latter ($p = .5$) influences the estimate half as much as the former ($p = 1$) under linear weighting (LW), a quarter under quadratic weighting (QW), and only an eighth under cubic weighting (CW). If no weighting method were used, then all responses would contribute to the estimate of treatment effect equally, regardless of reliability level.

The threshold method removed responses where the probability of honest responding (p) fell below the reliability level indicated by the threshold. Simulated thresholds ranged from 0.05 to 1.00 in increments of 0.05, with the additional inclusion of 0.99 (21 thresholds). As an example, for a threshold of 0.80, then only subjects with $p = 0.80$ or greater would be included in the analyses and everyone with $p < 0.80$ would be removed. Thresholds may be expressed either as a decimal or as a percentage (i.e., 0.80 or 80%).

Lastly, the combined method generated every possible combination of the weighting and threshold methods at each level for a total of 63 combinations. As an example, for the combination of 0.80 threshold with a quadratic weight, then all subjects with $p < 0.80$ would be removed from the data, then a quadratic weight would be imposed such that all remaining subjects would have their probability of responding reliably (p) squared (p^2) to create their individual weight.

Model Comparison

Organizational Structure

Each of the aforementioned bias correction strategies (87 in total) were incorporated into logistic regression models and were applied to each of the three unreliable response styles discussed above, resulting in 261 bias correction models. An additional four unadjusted models were generated: the "true" model and one for each of the three unreliable response styles, resulting in 265 models.

The 265 simulated models were generated for both sample reliability distributions (*n* = 2; ~50% and ~80%) resulting in 530 models. Further, these 530 models were implemented within each of the nine combinations of sample size (*n* = 3; *N* = 200, 2000, 20000) and treatment effect size (*n* = 3; *ES* = small, medium, large). Thus, a total of 4,770 models were generated for the current study. Consequently, the bias correction strategies are a function of sample size, effect size, reliability distribution, and unreliable response style.

The model estimates were organized first by sample size (*N*), then by treatment effect size (*ES*), resulting in a total of nine scenarios, as listed in the first row of Figure 1. Each of these nine scenarios contained both sample reliabilities (RLB), listed in the second row; each of the sample reliabilities contained all three unreliable response styles (RS) listed in the third row; and each of the unreliable response styles contained all three bias correction methods, listed in the bottom row. Thus, this figure is constructed in a top-down fashion such that each cell contains all cells in the subsequent rows. Herein, the nine scenarios listed in the top row are referred to as *N*/*ES* combinations and each sample reliability and response style combination (rows two and three) are referred to as RLB/RS combinations. Each *N*/*ES* combination contains all six RLB/RS combinations.

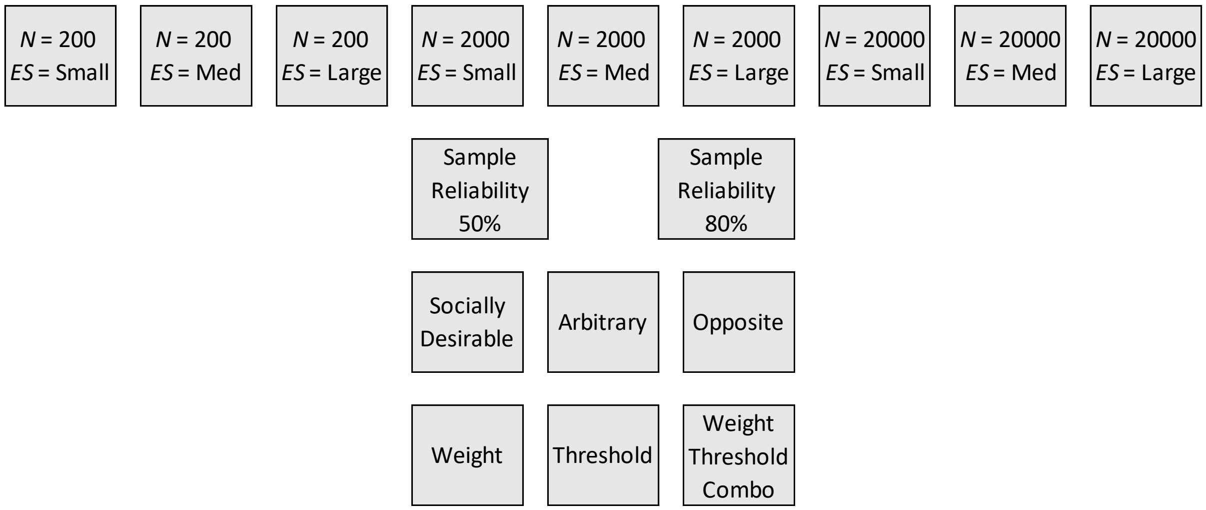

Figure 1: Organizational Structure of Bias Correction Models

***Bias* and *RRMSE* (%)**

Bias and *percent relative root mean square error* (*RRMSE* (%)) were calculated for the estimated treatment effect for each of the 4,770 models using the following formulas:

$$Bias = \left(\frac{1}{R}\sum_{i=1}^{R}\hat{\beta}_i\right) - \beta \qquad (4)$$

$$RRMSE\ (\%) = 100 \times \sqrt{\frac{1}{R}\sum_{i=1}^{R}\left(\hat{\beta}_i - \beta\right)^2} / |\beta| \qquad (5)$$

where R is the number of simulation replications, $\hat{\beta}_i$ is the estimated treatment effect size produced in the i^{th} replication, and β is the true treatment effect size. Furthermore, $\left(\frac{1}{R}\sum_{i=1}^{R}\hat{\beta}_i\right)$ is the sample equivalent of $E(\hat{\beta})$ (see formula 1), which is the expected value (i.e., mean) of the sampling distribution of the estimated treatment effect size generated from the full set of replications.

As indicated by formula (4), *bias* is the difference between the true treatment effect and the mean of the estimated treatment effects generated in a set of simulation replications. *Bias* is positive when the true value is overestimated and negative when underestimated. As indicated by formula (5), *RRMSE* (%) is the relative average distance between the true and individually estimated treatment effects and is calculated as relative rather than absolute in order to easily compare error across all models.

Although *bias* and *RRMSE* (%) were both calculated, only *RRMSE* (%) was used in selecting the bias correction strategies since it accounts for both *bias* and variance. Indeed, *RRMSE* (%) can be reduced to *MSE* (i.e., mean square error), which is equivalent to the variance plus *bias*-squared and is statistically decomposed as such:

$$RRMSE\ (\%) = 100\ \times \sqrt{\frac{1}{R}\ \sum_{i=1}^{R}\ (\ \hat{\beta}_i - \beta\ )^2} / \mid \beta \mid \quad (6)$$

$$MSE = \frac{1}{R}\sum_{i=1}^{R}\left[(\ \hat{\beta}_i - \beta\ )^2\right]$$

$$MSE = \frac{1}{R}\sum_{i=1}^{R}\left[\left[\left(\ \hat{\beta}_i - \frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i)\right) + \left(\frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i) - \beta\right)\right]^2\right]$$

$$MSE = \frac{1}{R}\sum_{i=1}^{R}\left[\left(\ \hat{\beta}_i - \frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i)\right)^2\right] + \frac{1}{R}\sum_{i=1}^{R}\left[\left(\frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i) - \beta\right)^2\right] + \frac{1}{R}\sum_{i=1}^{R}\left[2\left(\hat{\beta}_i - \frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i)\right)\left(\frac{1}{R}\sum_{i=1}^{R}(\hat{\beta}_i) - \beta\right)\right]$$

$$MSE = V(\hat{\beta}_i) + \left[B(\hat{\beta}_i)\right]^2$$

where $V(\hat{\beta}_i)$ is the variance and $\left[B(\hat{\beta}_i)\right]^2$ is the *bias*-squared.

Selecting the Best Strategy

Within each *N*/*ES* combination, all RLB/RS combinations were examined, and the model, other than the “true” model, that provided the treatment effect estimate with the lowest *RRMSE* (%) was chosen as the index bias correction strategy. Since the index strategies appeared in a random pattern across the RLB/RS combinations, which did not allow for general

Table 2: RRMSE (%) Margins by Scenario

N	*ES*	RLB	*RRMSE* (%)								
			Opposite			Arbitrary			Social Desirability		
			Min	Mgn	Max	Min	Mgn	Max	Min	Mgn	Max
200	S	50	93.453	3.5	96.953	87.375	3.5	90.875	98.997	3.5	102.497
		80	86.906	3.5	90.406	84.081	3.5	87.581	87.968	3.5	91.468
	M	50	46.085	2.5	48.585	43.132	2.5	45.632	44.537	2.5	47.037
		80	40.807	2.5	43.307	40.039	2.5	42.539	40.705	2.5	43.205
	L	50	34.609	2.0	36.609	32.764	2.0	34.764	51.434	2.0	53.434
		80	28.953	2.0	30.953	28.366	2.0	30.366	30.008	2.0	32.008
2000	S	50	33.699	2.0	35.699	31.548	2.0	33.548	31.360	2.0	33.360
		80	26.452	2.0	28.452	25.937	2.0	27.937	26.387	2.0	28.387
	M	50	16.305	1.5	17.805	15.427	1.5	16.927	14.516	1.5	16.016
		80	12.458	1.5	13.958	12.198	1.5	13.698	12.107	1.5	13.607
	L	50	12.075	1.5	13.575	11.669	1.5	13.169	10.655	1.5	12.155
		80	9.376	1.5	10.876	9.187	1.5	10.687	8.991	1.5	10.491
20000	S	50	11.781	1.0	12.781	11.353	1.0	12.353	10.042	1.0	11.042
		80	9.142	1.0	10.142	8.969	1.0	9.969	8.866	1.0	9.866
	M	50	5.614	0.5	6.114	5.482	0.5	5.982	4.877	0.5	5.377
		80	4.035	0.5	4.535	3.988	0.5	4.488	3.857	0.5	4.357
	L	50	3.990	0.5	4.490	3.968	0.5	4.468	3.468	0.5	3.968
		80	2.913	0.5	3.413	2.907	0.5	3.407	2.772	0.5	3.272

Note: N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large), RLB = Sample Reliability (%), Mgn = Margin.

applicability, an *RRMSE* (%) margin was generated such that all strategies producing estimates within the margin were considered for recommendation. In order to increase applicability of the bias correction strategies, the *RRMSE* (%) margin was increased in 0.5 increments until at least one strategy covered all six RLB/RS combinations. That is, all six RLB/RS combinations within each of the nine *N*/*ES* combinations contained treatment effect estiamtes that fell within the *RRMSE* (%) margin, allowing for general applicability. The *RRMSE* (%) margins used in each scenario are presented in Table 2. In cases where multiple strategies covered all six RLB/RS combinations, the strategy that generated the estimate with the lowest *RRMSE* (%) for the majority of RLB/RS combinations was chosen as the recommended strategy. This selection process resulted in a specific bias correction strategy recommendation for each of the nine *N*/*ES* combinations.

Results

Simulations were run in R version 3.3.2 and 1,000 replications were conducted to create the sampling distribution used to estimate the treatment effect size in each of the bias correction models, which are a function of sample size, effect size, reliability distribution, and unreliable response style.

Prediction Equations

Table 3 provides the prediction equations for the various combinations of treatment and covariate levels with the corresponding log odds and probability of outcome for each of these groups, by effect size. Baseline rates of the outcome are approximately 27% for group A and 44% for group B, with these rates decreasing after treatment to approximately 20% and 35%, respectively, for the small effect size, 13% and 24% for the medium effect size, and 8% and 16% for the large effect size.

Table 3: Prediction Equations, Log Odds of Outcome, and Probability of Outcome

Effect Size	Group	Prediction Equations	Log Odds of Outcome	Probability of Outcome
		$logit[P(Y=1)] = \beta_0 + \beta_1(treatment) + \beta_2(covariate)$		
None	Control, A	$logit[P(Y=1)] = \beta_0 + \beta_1(0) + \beta_2(0)$	-1.000	0.269
	Control, B	$logit[P(Y=1)] = \beta_0 + \beta_1(0) + \beta_2(1)$	-0.250	0.438
Small	Treat, A	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(0)$	-1.380	0.201
	Treat, B	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(1)$	-0.630	0.348
Medium	Treat, A	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(0)$	-1.915	0.128
	Treat, B	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(1)$	-1.165	0.238
Large	Treat, A	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(0)$	-2.420	0.082
	Treat, B	$logit[P(Y=1)] = \beta_0 + \beta_1(1) + \beta_2(1)$	-1.670	0.158

Note. A = group A; B = group B.

Subject- and Sample-Level Reliability

Figure 2 presents the reliability distributions for the *N* = 20000 samples with the top two graphs representing the entire sample and the bottom two graphs representing only individuals where $p < 1$. The 50% reliable sample is presented on the left and the 80% reliable sample on the right. The x-axis presents p and the y-axis presents the relative frequency (i.e., percentage

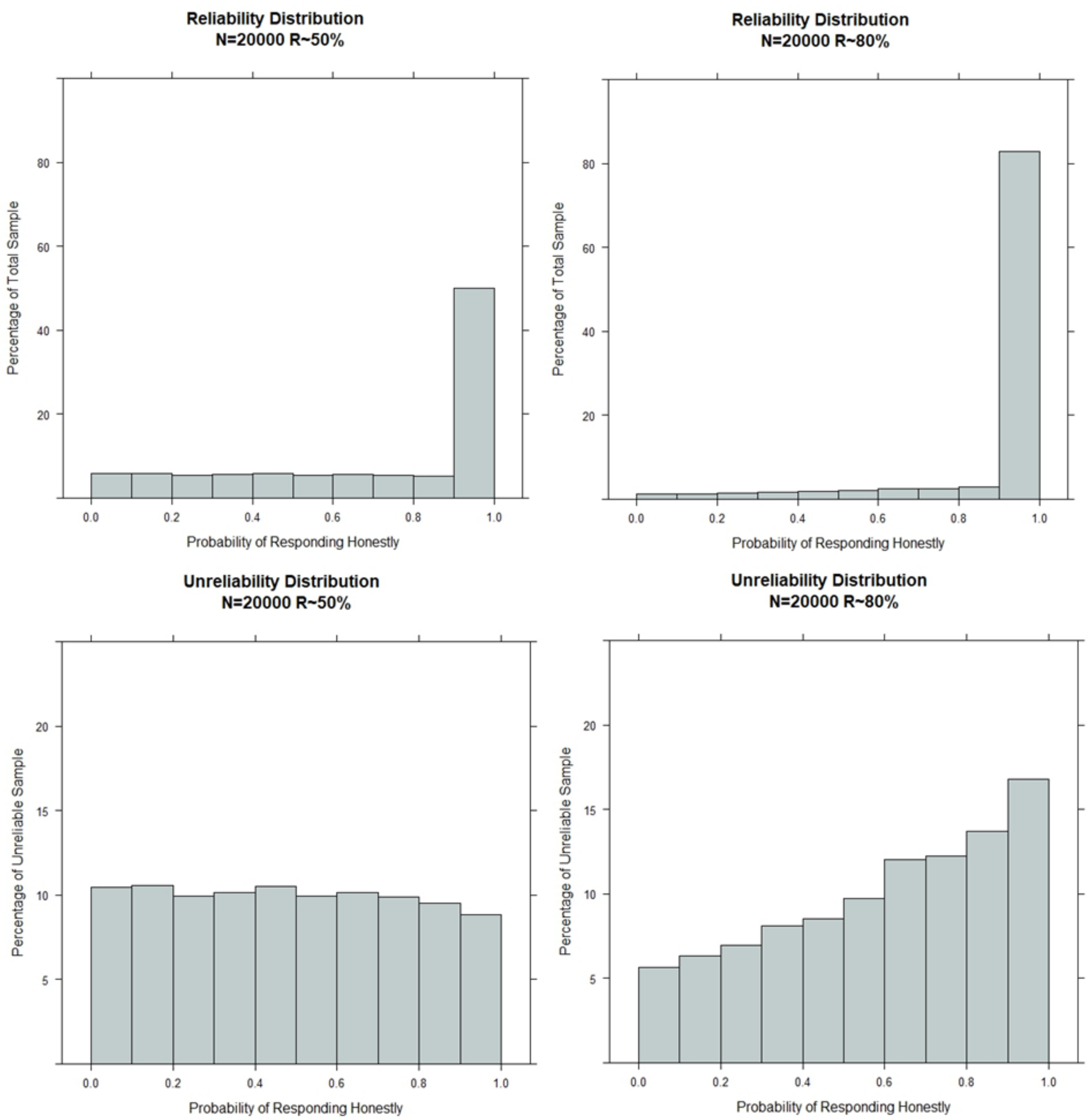

Figure 2: Reliability Distributions for 50% and 80% Reliable Samples (*N* = 20000)

of sample who falls in each category of p). Comparing the top graphs, the highest bar to the right represents the subjects where p = 1 (i.e., reliable responders). It is evident that this group makes up about 50% of the sample in the upper-left panel and about 80% in the upper-right. In these graphs, the lower bars to the left represent subjects where p < 1 (i.e., unreliable responders), which is highlighted in the lower graphs. Here, for the 50% reliable sample, each level of p contains approximately 10% of the unreliable sample, showing a minimal amount of variance among the different levels of unreliability. However, for the 80% reliable sample, the distribution of unreliability is skewed such that the percentage of subjects in each level of p increases as p increases.

Supplementing Figure 2, Table 4 presents descriptives of subject-level reliability for each sample reliability. Intuitively, the 80% reliable sample will have higher mean and median subject-level reliabilities since the overall reliability is higher, which is evidenced across all sample sizes. The mean subject-level reliability for the 50% reliable sample is approximately 70% whereas the mean for the 80% reliable sample is approximately 90%. Likewise, the median reliability for the 50% reliable sample is only around 90%, whereas it is consistently 100% in the 80% reliable sample.

Table 4: Subject-Level Reliability for 50% and 80% Reliable Samples

RLB	*N*	Min	Mean	Median	Max
50	200	0.003	0.712	0.884	1.000
	2000	0.001	0.714	0.893	1.000
	20000	0.000	0.719	0.898	1.000
80	200	0.063	0.898	1.000	1.000
	2000	0.001	0.913	1.000	1.000
	20000	0.000	0.917	1.000	1.000

Note. RLB = Sample Reliability (%).

Comparison of Bias Correction Strategies

Within each of the nine *N*/*ES* combinations, the index strategy was chosen for each RLB/RS combination, the *RRMSE* (%) margin was created to allow for increased applicability, and competing strategies that fell within the margin were compared. The bias correction strategy that produced estimates with the lowest *RRMSE* (%) for the majority of RLB/RS combinations was chosen as the best strategy for that *N*/*ES* combination.

Figures 3 through 5 present a comparison of select competing bias correction strategies within each of the nine *N/ES* combinations, separated by unreliable response style, with the graph for opposite responding appearing first, following by arbitrary responding, then socially desirable responding. The x-axis presents the 50% and 80% reliable samples and the y-axis is the *RRMSE* (%). The graphs are paneled by sample size, shown on the top horizontal side, and by treatment effect size, shown on the right vertical side.

The legend classifies levels of bias correction strategy by color and treatment effect size. Each color appears across all treatment effect sizes, but not necessarily across all sample sizes. The red points are the best bias correction strategies and the orange points represent the first alternate strategy, both of which appear across all sample sizes. The blue points represent the second alternate strategy and appear only for the *N* = 2000 and 20000 samples. Similarly, the green points represent the third alternate strategy and appear only for the *N* = 20000 sample. That is, *N* = 200 has one selected alternate strategy, *N* = 2000 has two alternates, and *N* = 20000 has three alternates.

The treatment effect size that each level of bias correction strategy within the legend applies to is indicated by the letter S (small), M (medium), or L (large) at the beginning of the

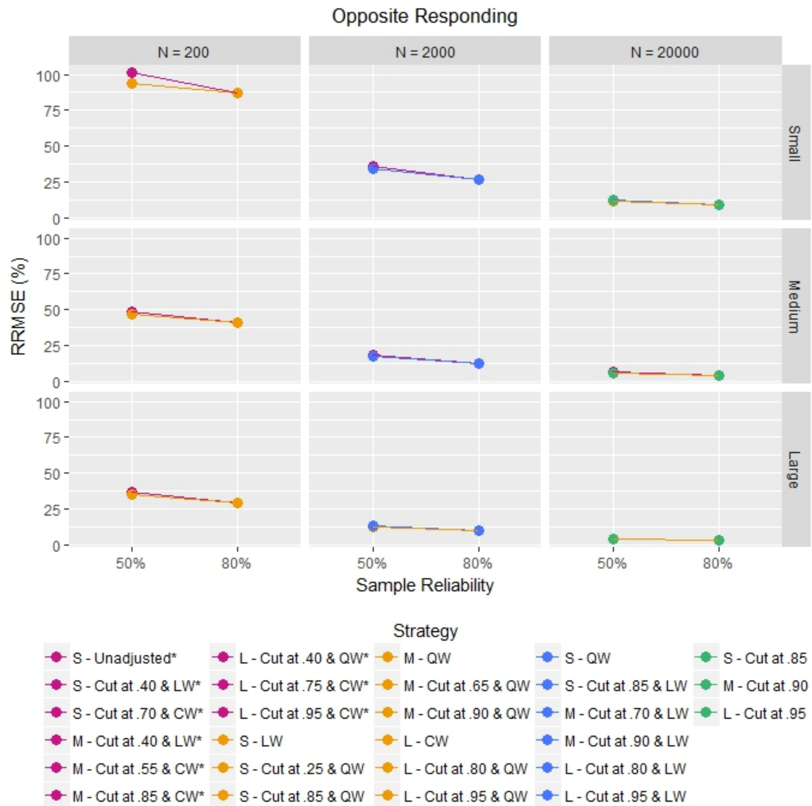

Figure 3: Comparison of *RRMSE (%)* for Competing Bias Correction Strategies, Opposite Responding

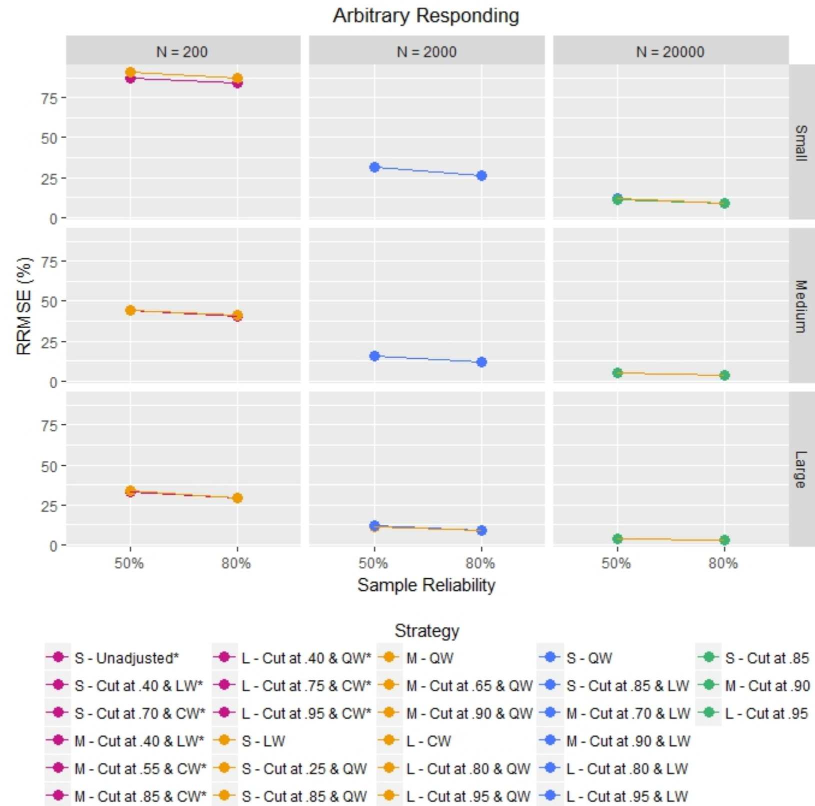

Figure 4: Comparison of *RRMSE (%)* for Competing Bias Correction Strategies, Arbitrary Responding

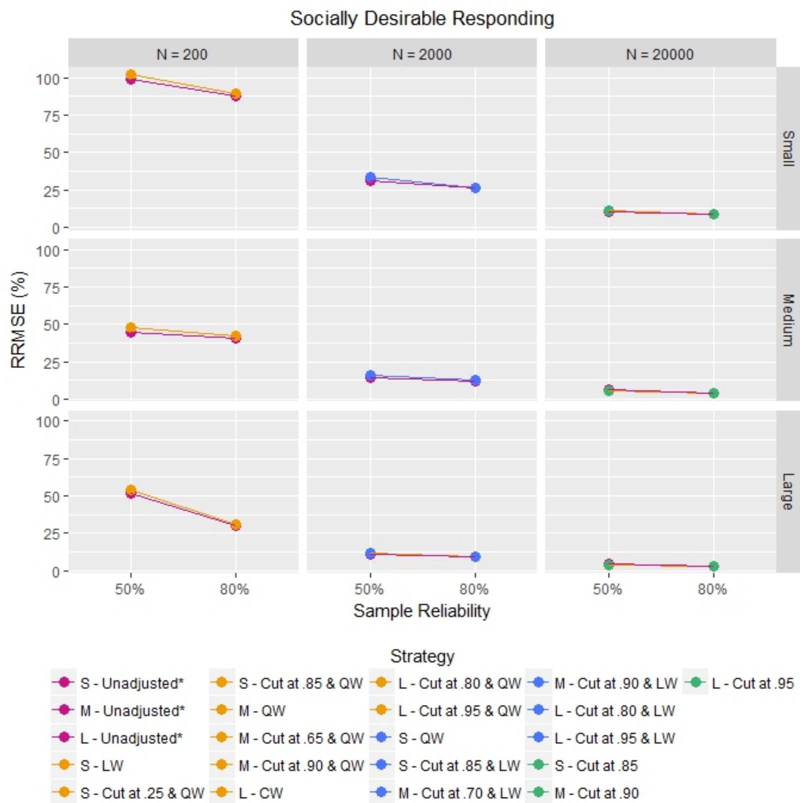

Figure 5: Comparison of *RRMSE (%)* for Competing Bias Correction Strategies, Socially Desirable Responding

level name. For opposite and arbitrary responding, the red points (best strategy) have three levels for each of the small, medium, and large effect sizes, one for each sample size, listed in order of appearance (*N* = 200, *N* = 2000, then *N* = 20000), indicating that each sample size had a different best strategy for each level of effect size. For social desirability, only one level of bias correction strategy is listed for each effect size since the same strategy was used across all sample sizes. For all three unreliable response styles, the orange points (first alternate) each have three levels of bias correction strategy for each effect size, with one for each sample size; the blue points each have two levels for each effect size, one for *N* = 2000 and the other for *N* = 20000; and the green points have only one level for each effect size, for *N* = 20000.

As evidenced in the graphs, there is not much variation among estimates from the competing bias correction strategies as they appear to lie directly upon one another with few exceptions, primarily for *N* = 200. Here, the red point (best strategy) falls below the orange point (alternate strategy) for most of the response styles in the 50% and 80% samples across all effect sizes. The exception to this is opposite responding in the 50% reliable sample, wherein the alternate strategy produces slightly less error; however, this effect diminishes in the 80% sample and is a small cost for widely increased applicability. Evaluating these graphs collectively, the significant amount of overlap in the points highlights similar error among the strategies, suggesting that the *RRMSE* (%) margin used to select the best bias correction strategy is reasonable as substantial error was not introduced in the process of increasing applicability.

To supplement Figure 3 and provide a more detailed comparison, Tables 5 through 7 present the *bias* and *RRMSE* (%) for the competing strategies for opposite, arbitrary, and socially desirable responding, respectively. These tables clearly show the number of alternate strategies that arose for each sample size and the strategies evaluated in each scenario. The sample size (*N*), effect size (*ES*), and sample reliabilities (RLB) are presented in the first three columns, defining each scenario, while the competing strategies are presented in the subsequent columns.

Examining Tables 5 and 6, for scenario *N* = 200/*ES* = small, the best strategy is no adjustment, as the one alternate strategy (LW) increases *RRMSE* (%) for most of the RLB/RS combinations, rather than decreasing it. The one exception is opposite responding in the 50% sample, but as previously mentioned, this is the cost of increasing applicability of the bias

Table 5: Bias and *RRMSE* (%) for Competing Bias Correction Strategies, Opposite Responding

N	*ES*	RLB	Best Strategy			Alternate 1			Alternate 2			Alternate 3		
			Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)
200	S	50	None	0.243	101.307	LW	0.125	93.453	------	------	------	------	------	------
		80	None	0.095	86.906	LW	0.044	87.128	------	------	------	------	------	------
	M	50	40%-LW	0.242	48.451	QW	0.205	46.546	------	------	------	------	------	------
		80	40%-LW	0.086	40.913	QW	0.065	41.038	------	------	------	------	------	------
	L	50	40%-QW	0.302	36.425	CW	0.233	34.851	------	------	------	------	------	------
		80	40%-QW	0.084	29.042	CW	0.054	29.238	------	------	------	------	------	------
2000	S	50	40%-LW	0.076	35.548	QW	0.063	34.505	25%-QW	0.060	34.277	------	------	------
		80	40%-LW	0.023	26.592	QW	0.017	26.453	25%-QW	0.016	26.452	------	------	------
	M	50	55%-CW	0.094	17.788	65%-QW	0.090	17.708	70%-LW	0.087	17.694	------	------	------
		80	55%-CW	0.032	12.544	65%-QW	0.032	12.576	70%-LW	0.032	12.635	------	------	------
	L	50	75%-CW	0.093	12.951	80%-LW	0.085	12.991	80%-QW	0.074	12.665	------	------	------
		80	75%-CW	0.037	9.436	80%-LW	0.035	9.510	80%-QW	0.031	9.441	------	------	------
20000	S	50	70%-CW	0.025	12.465	85%	0.013	11.866	85%-LW	0.012	11.819	85%-QW	0.011	11.792
		80	70%-CW	0.009	9.186	85%	0.005	9.161	85%-LW	0.005	9.150	85%-QW	0.004	9.144
	M	50	85%-CW	0.028	6.056	90%	0.018	5.698	90%-LW	0.017	5.667	90%-QW	0.016	5.641
		80	85%-CW	0.010	4.103	90%	0.007	4.062	90%-LW	0.006	4.055	90%-QW	0.006	4.049
	L	50	95%-CW	0.009	3.990	95%	0.010	4.003	95%-LW	0.010	3.999	95%-QW	0.010	3.994
		80	95%-CW	0.003	2.914	95%	0.003	2.913	95%-LW	0.003	2.913	95%-QW	0.003	2.913

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large), RLB = Sample Reliability (%), LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

Table 6: Bias and *RRMSE* (%) for Competing Bias Correction Strategies, Arbitrary Responding

N	*ES*	RLB	Best Strategy			Alternate 1			Alternate 2			Alternate 3		
			Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)
200	S	50	None	0.130	87.375	LW	0.064	90.87	------	------	------	------	------	------
		80	None	0.052	84.081	LW	0.024	86.862	------	------	------	------	------	------
	M	50	40%-LW	0.123	44.135	QW	0.101	44.012	------	------	------	------	------	------
		80	40%-LW	0.043	40.587	QW	0.032	40.983	------	------	------	------	------	------
	L	50	40%-QW	0.141	33.234	CW	0.098	33.751	------	------	------	------	------	------
		80	40%-QW	0.026	29.082	CW	0.010	29.480	------	------	------	------	------	------
2000	S	50	40%-LW	0.037	31.548	QW	0.030	31.767	25%-QW	0.028	31.821	------	------	------
		80	40%-LW	0.010	26.133	QW	0.007	26.201	25%-QW	0.007	26.225	------	------	------
	M	50	55%-CW	0.045	15.450	65%-QW	0.042	15.579	70%-LW	0.041	15.666	------	------	------
		80	55%-CW	0.015	12.236	65%-QW	0.015	12.237	70%-LW	0.016	12.264	------	------	------
	L	50	75%-CW	0.042	11.669	80%-LW	0.038	11.751	80%-QW	0.032	11.716	------	------	------
		80	75%-CW	0.018	9.216	80%-LW	0.018	9.271	80%-QW	0.015	9.266	------	------	------
20000	S	50	70%-CW	0.014	11.353	85%	0.008	11.618	85%-LW	0.007	11.621	85%-QW	0.007	11.635
		80	70%-CW	0.005	8.996	85%	0.003	9.080	85%-LW	0.003	9.081	85%-QW	0.002	9.086
	M	50	85%-CW	0.016	5.535	90%	0.011	5.501	90%-LW	0.010	5.492	90%-QW	0.010	5.486
		80	85%-CW	0.006	3.988	90%	0.004	4.007	90%-LW	0.004	4.006	90%-QW	0.003	4.006
	L	50	95%-CW	0.006	3.966	95%	0.006	3.973	95%-LW	0.006	3.970	95%-QW	0.006	3.968
		80	95%-CW	0.001	2.910	95%	0.001	2.907	95%-LW	0.001	2.908	95%-QW	0.001	2.909

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large), RLB = Sample Reliability (%), LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

Table 7: Bias and *RRMSE* (%) for Competing Bias Correction Strategies, Socially Desirable Responding

N	*ES*	RLB	Best Strategy			Alternate 1			Alternate 2			Alternate 3		
			Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)	Strategy	*Bias*	*RRMSE*(%)
200	S	50	None	-0.019	98.997	LW	-0.014	101.964	------	------	------	------	------	------
		80	None	-0.002	87.968	LW	0.000	89.337	------	------	------	------	------	------
	M	50	None	-0.006	44.537	QW	-0.016	47.866	------	------	------	------	------	------
		80	None	0.000	40.705	QW	-0.004	42.123	------	------	------	------	------	------
	L	50	None	-0.039	51.822	CW	-0.071	53.695	------	------	------	------	------	------
		80	None	-0.031	30.008	CW	-0.038	30.759	------	------	------	------	------	------
2000	S	50	None	0.033	31.360	QW	0.006	33.272	25%-QW	0.006	33.335	------	------	------
		80	None	0.012	26.439	QW	0.001	26.710	25%-QW	0.000	26.715	------	------	------
	M	50	None	0.046	14.623	65%-QW	0.000	15.615	70%-LW	-0.001	15.655	------	------	------
		80	None	0.018	12.195	65%-QW	0.000	12.379	70%-LW	0.000	12.423	------	------	------
	L	50	None	0.052	10.740	80%-LW	-0.008	11.780	80%-QW	-0.009	11.826	------	------	------
		80	None	0.021	9.053	80%-LW	0.000	9.304	80%-QW	0.000	9.312	------	------	------
20000	S	50	None	0.020	10.751	85%	0.003	11.529	85%-LW	0.003	11.553	85%-QW	0.003	11.584
		80	None	0.008	8.719	85%	0.001	9.058	85%-LW	0.001	9.065	85%-QW	0.001	9.073
	M	50	None	0.038	6.048	90%	0.003	5.394	90%-LW	0.003	5.398	90%-QW	0.003	5.402
		80	None	0.014	4.044	90%	0.001	3.995	90%-LW	0.001	3.997	90%-QW	0.001	3.999
	L	50	None	0.045	4.432	95%	0.002	3.959	95%-LW	0.002	3.958	95%-QW	0.002	3.957
		80	None	0.016	2.919	95%	0.000	2.908	95%-LW	0.000	2.909	95%-QW	0.000	2.910

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large), RLB = Sample Reliability (%), LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

Table 8: Bias and *RRMSE* (%) for Adjusted Social Desirability

N	*ES*	Strategy	50% Sample		80% Sample	
			Bias	*RRMSE*(%)	*Bias*	*RRMSE*(%)
200	S	None	------	------	------	------
	M	40%-LW	-0.013	46.895	-0.002	41.704
	L	40%-QW	-0.065	52.991	-0.036	30.534
2000	S	40%-LW	0.009	32.695	0.002	26.515
	M	55%-CW	0.000	15.526	0.000	12.392
	L	75%-CW	-0.008	11.726	0.000	9.293
20000	S	70%-CW	0.004	10.956	0.001	8.958
	M	85%-CW	0.004	5.307	0.001	3.963
	L	95%-CW	0.002	3.956	0.000	2.911

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large), LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

correction strategies. Reviewing the remainder of scenarios for opposite and arbitrary responding, results show that the best strategy produces estimes with lower *RRMSE* (%) than the alternative strategies.

Table 7 shows that the best strategy for social desirability in all scenarios is no adjustment. When compared to the unadjusted estimates, the best bias correction strategy in each scenario resulted in an increase in *RRMSE* (%), rather than a decrease, when applied to socially desirable responses. Thus, making no adjustment to these responses minimizes error in the estimator. However, to prevent introducing extra error into the estimate should the bias correction strategy be inadvertently applied to socially desirable responses, the *RRMSE* (%) for this response style was still taken into consideration when selecting the best strategy for each scenario. That is, if the best strategy in a scenario were applied to socially desirable responses, the *RRMSE* (%) for this response style would still fall within the acceptable margin, thus preventing a substantial increase in estimator error. Table 8 presents the *bias* and *RRMSE* (%) for social desirability if this were to happen.

Recommendations

The best bias correction strategies listed in Tables 5 and 6 are the recommended strategies and are summarized in Table 9 for each of the nine *N*/*ES* combinations. Overall, as sample and effect sizes increase, the bias correction strategy becomes more aggressive, as evidenced by the increasing reliability thresholds and weights.

Table 9: Recommended Bias Correction Strategies

N	*ES*	Strategy	
		Threshold (%)	Weight
200[a]	Small	None	None
	Medium	40	Linear
	Large	40	Quadratic
2000[a]	Small	40	Linear
	Medium	55	Cubic
	Large	75	Cubic
20000[a]	Small	70	Cubic
	Medium	85	Cubic
	Large	95	Cubic

Note. a – social desirability: no adjustment, regardless of sample/effect sizes.

For clinical trials with *N* = 200, when *ES* = small, no adjustment is the recommended strategy, but when *ES* = medium or *ES* = large, a 40% reliability threshold is recommended and should be combined with a linear or quadratic weight, respectively. For trials with *N* = 2000, when *ES* = small, a 40% threshold is recommended with a linear weight, but a cubic weight is recommended when *ES* = medium or *ES* = large with reliability thresholds of 55% and 75%, respectively. For trials with *N* = 20000, a cubic weight is recommended for all effect sizes, but reliability thresholds differ and are recommended at 70%, 85%, and 95% when *ES* = small, medium, and large, respectively.

Comparison of Unadjusted and Adjusted Estimates

Figures 6 though 8 present a comparison of the unadjusted and adjusted estimates within each of the nine *N*/*ES* combinations (i.e., study scenarios) for opposite, arbitrary, and socially desirable responding, respectively. Note that the adjusted estimates incorporate the recommended bias correction strategies for the given study scenario. The x-axis presents the 50% and 80% sample reliabilities and the y-axis is the *RRMSE* (%). The graphs are paneled by sample size, shown on the top horizontal side, and by treatment effect size, shown on the right vertical side. The red points represent the recommended bias correction strategies presented above and the orange points represent the unadjusted models. For scenarios where no adjustment is the recommended strategy (i.e., *N* = 200/*ES* = small and socially desirable responding) the adjusted and unadjusted models are one and the same, thus, there is no distinction between the red and orange points in these cases.

Examining *RRMSE* (%) for the unadjusted models (orange points) across all three graphs, it is apparent that responding styles that introduce more *bias* into the model have higher error, just as expected, with this effect more pronounced in the 50% reliable samples. Comparing across the nine *N*/*ES* combinations, unadjusted *RRMSE* (%) remains relatively consistent across sample and effect sizes within each responding style. This trend is indicated by the orange points appearing in approximately the same location within each study scenario, although a little more variability is seen for the *N* = 200 trials compared to the other sample sizes, reflecting increased variation due to the small sample size.

Inspecting *RRMSE* (%) for the recommended bias correction strategies (red points), results show that, within each study scenario, the amount of adjusted error is similar in both

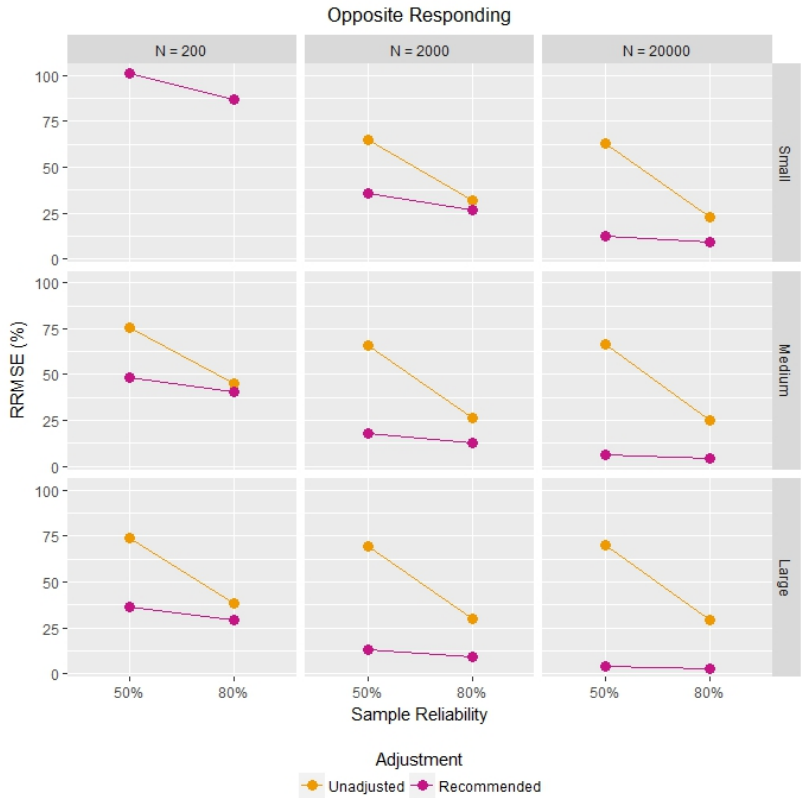

Figure 6: Comparison of *RRMSE (%)* for Unadjusted and Adjusted Models, Opposite Responding

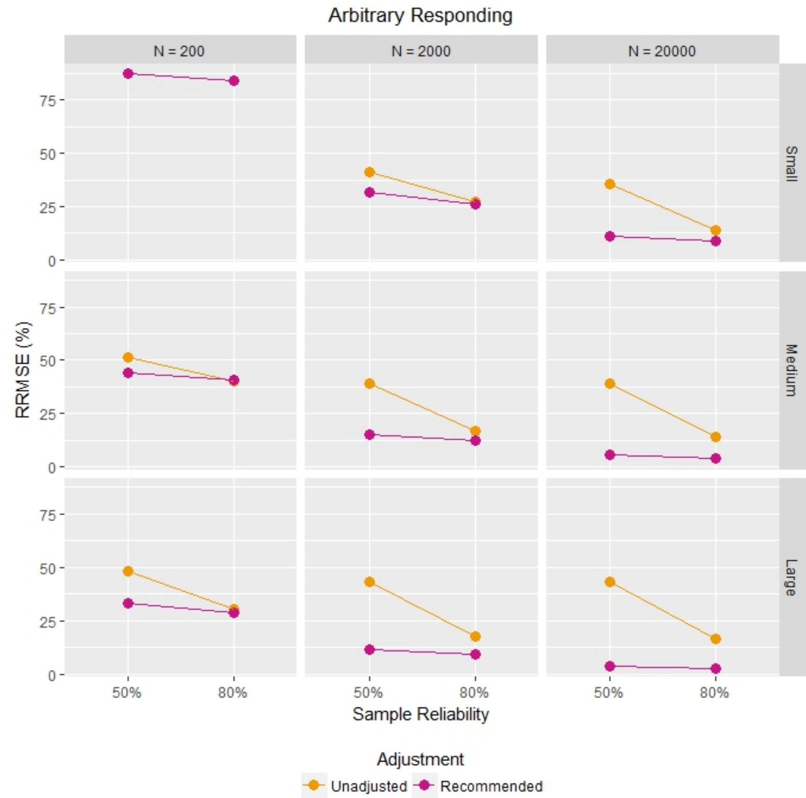

Figure 7: Comparison of *RRMSE (%)* for Unadjusted and Adjusted Models, Arbitrary Responding

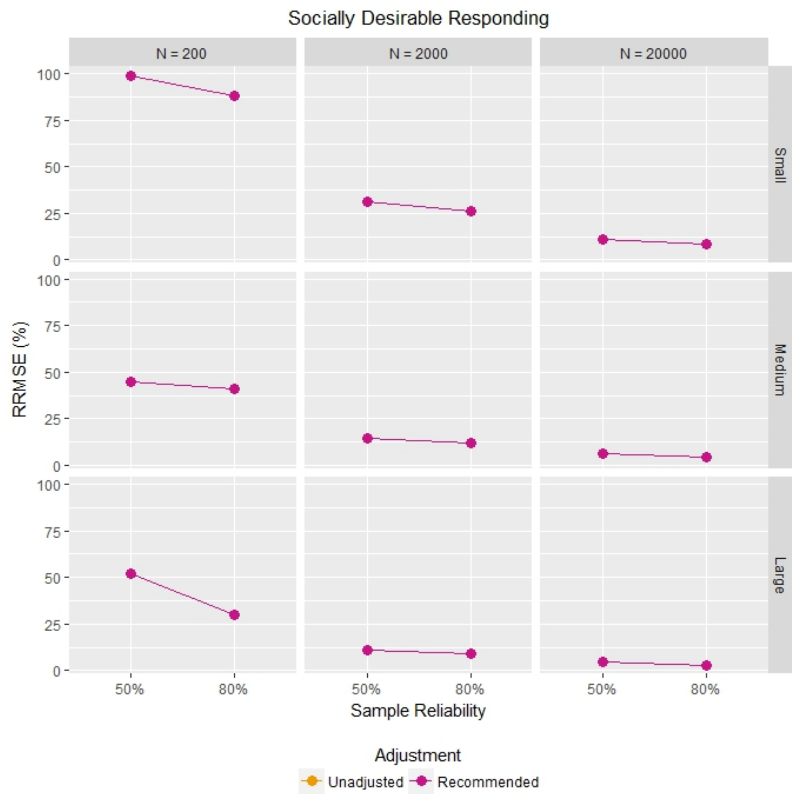

Figure 8: Comparison of *RRMSE (%)* for Unadjusted and Adjusted Models, Socially Desirable Responding

sample reliabilities, as indicated by the relative lack of slope in the red lines. However, steeper slopes do emerge for trials with *N* = 200 due to increased variance. Adjusted error is also similar across responding styles for a given study scenario, as indicated by the red points appearing in approximately the same location in each of their respective plots. Comparing across the nine *N*/*ES* combinations, *RRMSE* (%) decreases as both sample and effect sizes increase, evidenced by the red lines appearing closer and closer to the bottom of each plot as the graph progress from left to right in each row and from top to bottom in each column.

The reduction in *RRMSE* (%) in the bias corrected estimates as sample and effect sizes increase, in conjunction with the consistency of the *RRMSE* (%) in the unadjusted estimates, suggests that the magnitude of the bias correction increases as sample and effect sizes increase. The magnitude is also larger for the less reliable scenarios; that is, responding styles that introduce higher *bias* and for the 50% reliable samples. Indeed, these magnitudes are evidenced by the increasing distance between the orange and red lines as sample and effect sizes increase (moving left to right across each row and from top to bottom in each colulmn) and as responding style moves from less biased (social desirability) to more biased (opposite), with these trends especially pronounced in the 50% reliable sample.

Tables 10 and 11 present the *bias* and *RRMSE* (%) for estimates from the unadjusted and adjusted models for opposite and arbitrary responding, respectively, as well as the amount of *bias* and *RRMSE* (%) reduction, and the strategy used in the bias correction. When no adjustment was the recommended strategy, dashed lines were entered for the adjusted and reduction columns. For socially desiarability, Table 12 presents the *bias* and *RRMSE* (%) for estimates from the unadjusted models only since no bias corrections were applied.

Although *RRMSE* (%) increased by 0.296 for arbitrary responding in the 80% reliable *N* = 200/*ES* = medium trial (Table 11, row 4), the increase is negligible and is a small price to pay for decreased error for the other RLB/RS combinations in this scenario (i.e., ranging from 4.357 to 26.828). Other than that, all *RRMSE* (%) changes are reductions ranging from 1.430 to 39.562 for arbitrary responding and from 4.357 to 65.937 for opposite responding across both 50% and 80% reliable samples. These numbers correspond to the distance between the red and orange points in Figures 6 and 7, again showing that opposite responding experiences the largest reduction in error, followed by arbitrary responding, with the same trend seen for *bias*. Tables 10 and 11 also show that *bias* and *RRMSE* (%) decrease as sample and effect sizes increase.

Table 10: Bias and *RRMSE* (%) for Unadjusted and Adjusted Models, Opposite Responding

N	*ES*	RS	Unadjusted		Adjusted		Reduction		Strategy
			Bias	*RRMSE*(%)	*Bias*	*RRMSE*(%)	*Bias*	*RRMSE*(%)	
200	S	50	0.243	101.307	------	------	------	------	None
		80	0.095	86.906	------	------	------	------	None
	M	50	0.619	75.279	0.242	48.451	0.377	26.828	40%-LW
		80	0.242	45.270	0.086	40.913	0.156	4.357	40%-LW
	L	50	1.001	73.687	0.302	36.425	0.699	37.262	40%-QW
		80	0.413	38.250	0.084	29.042	0.329	9.208	40%-QW
2000	S	50	0.228	64.763	0.076	35.548	0.152	29.215	40%-LW
		80	0.074	31.698	0.023	26.592	0.051	5.106	40%-LW
	M	50	0.595	65.865	0.094	17.788	0.501	48.077	55%-CW
		80	0.221	26.547	0.032	12.544	0.189	14.003	55%-CW
	L	50	0.983	69.528	0.093	12.951	0.890	56.577	75%-CW
		80	0.407	29.690	0.037	9.436	0.370	20.254	75%-CW
20000	S	50	0.237	62.777	0.025	12.465	0.212	50.312	70%-CW
		80	0.082	23.001	0.009	9.186	0.073	13.815	70%-CW
	M	50	0.606	66.260	0.028	6.056	0.578	60.204	85%-CW
		80	0.229	25.265	0.010	4.103	0.219	21.162	85%-CW
	L	50	0.993	69.927	0.009	3.990	0.984	65.937	95%-CW
		80	0.415	29.337	0.003	2.914	0.412	26.423	95%-CW

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large),
LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

Table 11: Bias and *RRMSE* (%) for Unadjusted and Adjusted Models, Arbitrary Responding

N	*ES*	RS	Unadjusted		Adjusted		Reduction		Strategy
			Bias	*RRMSE*(%)	*Bias*	*RRMSE*(%)	*Bias*	*RRMSE*(%)	
200	S	50	0.130	87.375	------	------	------	------	None
		80	0.052	84.081	------	------	------	------	None
	M	50	0.353	51.462	0.123	44.135	0.230	7.327	40%-LW
		80	0.130	40.291	0.043	40.587	0.087	-0.296	40%-LW
	L	50	0.607	48.550	0.141	33.234	0.466	15.316	40%-QW
		80	0.214	30.512	0.026	29.082	0.188	1.430	40%-QW
2000	S	50	0.121	40.910	0.037	31.548	0.084	9.362	40%-LW
		80	0.038	27.090	0.010	26.133	0.028	0.957	40%-LW
	M	50	0.344	39.109	0.045	15.450	0.299	23.659	55%-CW
		80	0.118	17.191	0.015	12.236	0.103	4.955	55%-CW
	L	50	0.607	43.357	0.042	11.669	0.565	31.688	75%-CW
		80	0.227	17.895	0.018	9.216	0.209	8.679	75%-CW
20000	S	50	0.132	35.602	0.014	11.353	0.118	24.249	70%-CW
		80	0.044	14.027	0.005	8.996	0.039	5.031	70%-CW
	M	50	0.355	38.976	0.016	5.535	0.339	33.441	85%-CW
		80	0.124	13.941	0.006	3.988	0.118	9.953	85%-CW
	L	50	0.617	43.528	0.006	3.966	0.611	39.562	95%-CW
		80	0.231	16.418	0.001	2.910	0.230	13.508	95%-CW

Note. N = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large),
LW = Linear Weight, QW = Quadratic Weight, CW = Cubic Weight.

Table 12: Bias and *RRMSE* (%) for Unadjusted and Adjusted Models, Socially Desirable Responding

N	*ES*	50% Sample		80% Sample	
		Bias	*RRMSE*(%)	*Bias*	*RRMSE*(%)
200	S	-0.019	98.997	-0.002	87.968
	M	-0.006	44.537	0.000	40.705
	L	-0.039	51.822	-0.031	30.008
2000	S	0.033	31.360	0.012	26.439
	M	0.046	14.623	0.018	12.195
	L	0.052	10.740	0.021	9.053
20000	S	0.020	10.751	0.008	8.719
	M	0.038	6.048	0.014	4.044
	L	0.045	4.432	0.016	2.919

Note. *N* = Sample Size, *ES* = Effect Size (S = Small, M = Medium, L = Large).

Conclusion

Response bias is a common issue in survey research with well-documented impacts on statistical estimates. Although some attempts have been made at correcting for this bias, none are applicable when inconsistent reporting is the primary method of detecting and quantifying response bias. To address this gap, the current study analyzed 4,770 logistic regression models to evaluatae a wide variety of bias correction strategies across a myriad of scenarios (various combinations of sample size, effect size, sample reliability, and response style) to examine their collective impacts on estimating the treatment effect in a series of hypothetical clinical trials.

The simulations showed that as sample size increased, the number of available bias correction strategies also increased, with *N* = 200 having only one alternative strategy, *N* = 2000 having two, and *N* = 20000 having three. Additionally, *bias* and *RRMSE* (%) became more similar among estimates from competing strategies as sample and effect sizes increased, providing evidence of increasing estimator stability.

Based on these simulations, no adjustment is recommended for socially desirable responding or when *N* = 200/*ES* = small. However, when *N* = 200 and *ES* = medium or large, a 40% reliability threshold combined with a linear weight for *ES* = medium or a quadratic weight for *ES* = large is recommended. For *N* = 2000 when *ES* = small, a 40% threshold in combination with a linear weight is recommended, but a cubic weight is recommended when *ES* = medium and *ES* = large with reliability thresholds of 55% and 75%, respectively. For *N* = 20000, a cubic weight is recommended for all effect sizes, but reliability thresholds differ and are recommended at 70%, 85%, and 95% when *ES* = small, medium, and large, respectively.

Employing these bias correction strategies when unreliable responding has been detected within a dataset will decrease error and increase the accuracy of estimates and validity of inferences. In the current study, the best strategy for socially desirable responding was no adjustment since this responding style only introduced a small amount of bias and made minimal changes to the true responses. Thus, applying these strategies to responding styles that introduce minimal bias is not recommended. However, inadvertenly doing so will not introduce substantial error to the model and resulting estimates will still be valid. The results of this study also showed significant improvements in error for arbitrary and opposite responding, which introduced moderate and large amounts of bias into the model, respectively. Thus, applying these strategies to responding styles that introduce similar amounts of error will likely result in similar reductions in error. Furthermore, samples with low levels of reliability will receive the greatest benefits from these bias correction models; however, these strategies were designed to be applied to samples with higher levels of reliability as well, wherein more aggressive bias correction strategies can be applied.

Chapter 4: Discussion

The objective of this study is to provide a means of reducing error in treatment effect estimation during statistical modeling and to show how this small advancement fits into the big picture of bias in clinical and epidemiological research. Bias is a problem that has plagued researchers for decades and comes in many different forms and from many different angles. Whatever the form, bias introduces error into statistical modeling and estimation, decreasing the validity of findings and inferences, with the potential to mislead the scientific community or the general public.

For example, the optimal dosage of a drug could be inaccurately estimated in a clinical trial, resulting in an ineffective intervention if the dosage is underestimated or toxicity if overestimated. Furthermore, resources, namely grant funding, may not be distributed in a way that produces the best outcomes if the true effects are not evident due to biased results and/or reporting.

Response bias is a particularly problematic type of bias because there are many subtypes, but prevention efforts are typically aimed at social desirability while ignoring the other subtypes, making this a bias typically handled in the analysis phase of a study. One of the more common methods of detecting and quantifying response bias after a study has ended is by identifying inconsistent reporting in survey responses. However, current bias correction methods are not appropriate for use when such is the case, thus motivating the current study.

The recommended bias correction strategies presented in Chapter 3 require three pieces of information before they can be applied: (1) sample size, (2) effect size, and (3) subject-level reliability. The sample size will already by known and the effect size can be obtained by fitting the model as usual (i.e., unadjusted). Once these first two pieces of information are obtained, the third can be calculated as discussed below.

Subject-level reliability may be obtained either through historical data or by identifying inconsistent reporting within a given dataset. For the latter, the first step is to count the number of times that a subject had the opportunity to provide an illogically inconsistent response. This can occur across question sets (e.g., responding “No” to “Have you ever had sex?” but “Yes” to “Have you ever been pregnant?”), time points (e.g., responding “Yes” to “Have you ever had sex?” at baseline, but “No” at follow-up), or data collection methods (e.g., responding “No” to “Have you ever had sex?” in an interview, but blood or urine sample shows that subject is pregnant). Once the number of opportunities for inconsistent reporting has been identified, the next step is to calculate the number of times that each subject provided an inconsistent response. Then calculate the proportion of inconsistent responses for each subject, which is found by dividing the number of observed inconsistent responses by the number of opportunities to provide an inconsistent response. For example, if there are 10 sets of questions where inconsistent reporting is possible, and a subject provides inconsistent reports in 2 of these question sets, then the proportion of inconsistent responses would be $\frac{2}{10}$ = 0.20 or 20%. This number provides the probability of responding unreliably, so the inverse would be the probability of responding reliably (p) – which is what we are after – obtained by subtracting the probability of responding unreliably from 1. In this example, p = 1 – 0.20 = 0.80 or 80%.

Once all three pieces of required information are obtained, researchers can refer to Table 9, where the recommended bias correction strategies are presented, and select the strategy that applies to the sample and effect size combination relevant to their trial. If the sample size falls between those presented in Table 9, researchers are encouraged to modify the threshold component by the amount equivalent to the distance between the sample size in question and those presented in the table. For example, a sample size of 1000 with a small effect size is approximately half way between 200 and 2000, so the researcher can reduce the threshold from 40% to 20% and try either linear or no weighting.

Table 13: Step-by-Step Guide to Applying Bias Correction Strategies

Step	Instruction
1	Obtain sample size
2	Fit model as usual, obtain effect size
3	Count number of opportunities for inconsistent reporting
4	Calculate number of times each subject provided inconsistent response
5	Calculate proportion of inconsistent responses for each subject
6	Calculate 1 - [value from Step 6] to obtain probability of responding honestly
7	Select bias correction strategy that applies to sample and effect size
8	Refit model, applying threshold and weight
9	Compare standard error for unadjusted model from Step 2 and adjusted model from Step 9
10	Retain estimate from model that produced the smallest standard error

Assuming the recommended strategy is not “no adjustment,” the next step is to refit the model, first applying the subset analysis (i.e., threshold), then entering the weight into the model fitting statemennt to be applied to the retained portion of the sample. After applying the bias correction strategy, the standard error of the estimate from the unadjusted (first model) should be compared to that of the adjusted (second model) to verify validity of estimates and ensure the appropriate adjustment was applied. The estimate with the lower standard error should be retained and reported. See Table 13 for a summary of the step-by-step process to applying the bias correction strategies.

Limitations

When applying the bias correction strategies recommended in this study, researchers should consider the impact of these strategies on statistical power, which is the probability of correctly rejecting the null hypothesis and depends greatly upon sample size, effect size, and significance level. In applying these bias correction strategies, the threshold component directly reduces the sample size by removing subjects with a probability of responding honestly that is lower than the recommended threshold, which results in a reduction in power. Similarly, the weighting component reduces the influence of less-reliable responses on the estimate, indirectly decreasing the sample size since less information is contributed to the model. Thus, it is recommended that researchers conduct a power analysis both before and after the utilization of these bias correctio strategies to evaluate both the improvement in error and the potential reduction in power when considering the application of these strategies.

In the design phase of a study, when researchers are looking to other studies for estimates of effect size and attrition rates, they should also look for information on reliability distributions in similar samples. Researchers should then use that information in conjunction with sample size and effect size to predict which bias correction strategy will be most appropriate for their study. Doing so will allow researchers to account for the reduction in sample size that results from applying the tentative bias correction strategy. That is, just as sample sizes are increased to account for dropouts, they should also be increased to account for the proportion of the sample that will be removed when applying these bias correction strategies, as doing so will help maintain sufficient power. When reporting results, researchers should include their sample reliability distributions to aid other researchers in determining their own sample size calculations. Sample-level reliability can be reported as the proportion of sample that is 100% reliable (no inconsistent reports), or even 95% or 99% reliable, or as the mean of the subject-level reliability. Providing this information to assist other researchers will strengthen the integrity of the work produced in the scientific community.

Unfortunately, response bias often goes undetected and bias corrections are only as helpful as unreliability is detectable. In the context of a binary outcome, for example "Have you ever had sex before?" asked at two time points, the only options for true responses are *No*/*No, No*/*Yes,* and *Yes*/*Yes*. Under opposite responding, self-reports would emerge as *Yes*/*Yes, Yes*/*No,* or *No*/*No,* respectively, with the *Yes*/*No* response being the only one that is detectable (33.3% of unreliable responses). Under arbitrary responding, self-reports could appear as *No*/*No, No*/*Yes, Yes*/*No,* or *Yes*/*Yes*, again with the *Yes*/*No* response being the only one that is detectable (25% of unreliable responses). Under social desirability, assuming the social norm is not having had sex before, the self-reported response would be *No*/*No*, which is undetectable (0% of unreliable responses). Even when we are able to detect inconsistent reporting, there is no way to know which answer is true and which is not, or if neither are true since there is no way to truly know which responding style the subject is using.

Directions for Future Research

Although the current study has made advances in correcting for response bias in clinical trails where inconsistent reporting is the method used to identify response bias, only a limited number of scenarios were investigated. Future research would benefit from exploring how other scenarios would influence the choice of bias correction strategy as well. For example, the current study only utilized main effects models, but exploring these strategies in the context of interactions, especially those between the treatment effect and other covariates, would be particularly beneficial. Exploring the behavior of these strategies using more complex models like mixed effect models or generalized estimating equations (GEE) would also be helpful in increasing the applicability of these strategies.

Future research should also examine different levels of the variables that were included in the current study (i.e., sample size, effect size, reliability distribution, response style, and reliability weight) to see how other levels would influence the best bias correction strategy for various scenarios. For example, many studies have samples as small as 50 or 100 subjects, or as large as 50,000 or 100,000 or more so these other levels are important to explore.

Although this study covers three categories of effect size (i.e., small, medium, large), an odds ratio (*OR*) to Cohen's *d* effect size conversion was used (see Chen et al., 2010), which only provided conversions in scenarios where the rate of the outcome occurred in 1% - 10% of the unexposed/control group. However, the outcome was prevalent in 27% - 44% of the sample in the current study, which is out of the range of conversion. As 10% was the highest prevalence rate converted in the paper, the *OR*s corresponding to that conversion were used in the simulations. Reviewing Table 1 in Chen et al. (2010), *OR*s within each Cohen's *d* effect size category decrease as the prevalence rate increases. The *OR*s for a 10% prevalence rate are 1.46, 2.50, and 4.14 for small, medium, and large effect sizes, respectively, which are the values used in the current study (more precisely, their logarithmic transformed inverses). Since the prevalence rate in this study is 3 to 4 times larger than that used in the conversions, the *OR*s may actually apply to different effect size strengths. That is, 1.46 could potentially be considered a medium effect size and 2.50 a large effect size with prevalence rates as high as those in the current study. Without conversions corresponding to higher prevalence rates, the best option was to use the conversion for the highest rate available, 10%. Future research would benefit from examining prevalence rates of binary outcomes that have a direct conversion (i.e., 1% - 10%) as well as from extending the *OR* to Cohen's *d* conversions to include higher prevalence rates.

Regarding reliability distributions, the current study covers samples with reliabilities between approximately 50% and 80%, but other sample reliabilities may be higher or lower and should thus be explored. For instance, which bias correction strategies would be best for scenarios where reliability is closer to 30% or 40%? 90% or 95%? At which level of sample reliability will applying the strategies become futile because the reduction in error is so small that it is no longer worth the computational and application efforts? Also, this study calculates the sample-level reliability based on the proportion of subjects with a probability of responding reliably (p) of 100%, but future simulation studies may consider using (p) of 95% or 99% as well, or even using the mean of the subject-level reliability.

The present study used the same unreliable response style in both self-report variables within the models (i.e., outcome and covariate), but it may be advantageous to mix and match response styles within a single model such that one variable is impacted by social desirability responding and the other by careless responding (or any other combination). Other unreliable response styles should also be considered so that greater applicability of bias corrections can be obtained. For example, acquiescence bias and extreme responding are other types of response bias that occur in survey research but were not investigated in the current study.

Linear, quadratic, and cubic weights were used in this study, but future research would benefit from examining higher order weights, as these may further reduce estimator error. Indeed, reviewing Table 9, weights progress from none to cubic through the first half of the recommendations, but then remain at cubic weight for the second half. The inclusion of higher order weights would likely show a continued trend of increasing weight as the sample and effect sizes increase.

Countless different scenarios occur in research settings and the closer researchers come to mimicking these scenarios in the search for the best bias correction strategies, the more accurate our estimates and inferences will become. Although this study provides a foundation for applying these strategies, it is up to us as a scientific community to expand these findings, refining when and where these strategies are best applied to increase their general applicability and to improve the estimates and inferences we generate.

References

Adams, A. S., Soumerai, S. B., Lomas, J., & Ross-Degnan, D. (1999). Evidence of self-report bias in assessing adherence to guidelines. *International Journal for Quality in Health Care, 11*(3), 187-192. doi:10.1093/intqhc/11.3.187

Aguinis, H., Pierce, C. A., & Quigley, B. M. (1993). Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking: A meta-analytic review. *Journal of Applied Social Psychology, 23*(5), 352-373. doi:10.1111/j.1559-1816.1993.tb01092.x

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K., & Yeatts, K. B. (2015a). Selection Bias. *ERIC Notebook, 2*(13), 1-3. Retrieved from https://sph.unc.edu/files/2015/07/nciph_ERIC13.pdf

Alexander, L. K., Lopes, B., Ricchetti-Masterson, K., & Yeatts, K. B. (2015b). Sources of systematic error or bias: Information bias. *ERIC Notebook, 2*(14), 1-5. Retrieved from https://sph.unc.edu/files/2015/07/nciph_ERIC14.pdf

Alexander, M. G., & Fisher, T. D. (2003). Truth and consequences: using the bogus pipeline to examine sex differences in self-reported sexuality. *Journal of Sex Research, 40*(1), 27-35. doi:10.1080/00224490309552164

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare, 9*, 211-217. doi:10.2147/JMDH.S104807

Batchelor, J. H., & Miao, C. (2016). Extreme response style: A meta-analysis. *Journal of Organizational Psychology, 16*(2), 51-62. Retrieved from https://search.proquest.com/openview/9b9a4e2fdddadbd4bd4e3d323d450dc2/1?pq-origsite=gscholar&cbl=1576346

Beaussart, M. L., & Kaufman, J. C. (2013). Gender differences and the effects of perceived internet privacy on self-reports of sexual behavior and sociosexuality. *Computers in Human Behavior, 29*(6), 2524-2529. doi:10.1016/j.chb.2013.06.014

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48*(3), 386-398. doi:10.2307/2095230

Braga, L. H., Farrokhyar, F., & Bhandari, M. (2012). Confounding: What is it and how do we deal with it? *Canadian Journal of Surgery, 55*(2), 132-138. doi:10.1503/cjs.036311

Brault, M. C., Turcotte, O., Aimé, A., Côté, M., & Bégin, C. (2015). Body mass index accuracy in preadolescents: Can we trust self-report or should we seek parent report? *Journal of Pediatratics, 167*(2), 366-371. doi:10.1016/j.jpeds.2015.04.043

Brener, N. D., Grunbaum, J. A., Kann, L., McManus, T., & Ross, J. (2004). Assessing health risk behaviors among adolescents: The effect of question wording and appeals for honesty. *Journal of Adolescent Health, 35*(2), 91-100. doi:10.1016/j.jadohealth.2003.08.013

Brinker, G. D. (2002). Using standard scores to control for extreme response style bias. *Journal of Applied Sociology, 19*(2), 81-99. Retrieved from http://www.jstor.org/stable/43481457?seq=1#page_scan_tab_contents

Cabooter, E., Weijters, B., Geuens, M., & Vermeir, I. (2016). Scale format effects on response option interpretation and use. *Journal of Business Research, 69*(7), 2574-2584. doi:10.1016/j.jbusres.2015.10.138

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860-864. doi:10.1080/03610911003650383

Clarke, P., Sastry, N., Duffy, D., & Ailshire, J. (2014). Accuracy of self-reported versus measured weight over adolescence and young adulthood: Findings from the National Longitudinal Study of Adolescent Health, 1996-2008. *American Journal of Epidemiology, 180*(2), 153-159. doi:10.1093/aje/kwu133

Cox, E., Martin, B. C., Van Staa, T., Garbe, E., Siebert, U., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: Approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report--Part II. *Value Health, 12*(8), 1053-1061. doi:10.1111/j.1524-4733.2009.00601.x

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology, 24*, 349-354. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13813058

D'Agostino, R. B., Jr. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*(19), 2265-2281. doi:10.1002/(SICI)1097-0258(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B

Delgado-Rodriguez, M., & Llorca, J. (2004). Bias. *Journal of Epidemiology and Community Health, 58*(8), 635-641. doi:10.1136/jech.2003.008466

Dodou, D., & Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*(C), 487-495. doi:10.1016/j.chb.2014.04.005

Elgar, F. J., Roberts, C., Tudor-Smith, C., & Moore, L. (2005). Validity of self-reported height and weight and predictors of bias in adolescents. *Journal of Adolescent Health, 37*(5), 371-375. doi:10.1016/j.jadohealth.2004.07.014

Fadnes, L. T., Taube, A., & Tylleskar, T. (2009). How to identify information bias due to self-reporting in epidemiological research. *The Internet Journal of Epidemiology, 7*(2). Retrieved from http://ispub.com/IJE/7/2/9194

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35*(3), 263-282. doi:10.1177/0022022104264122

Giraudeau, B., & Ravaud, P. (2009). Preventing bias in cluster randomised trials. *PLoS Medicine, 6*(5), e1000065. doi:10.1371/journal.pmed.1000065

Gorber, S. C., Shields, M., Tremblay, M. S., & McDowell, I. (2008). The feasibility of establishing correction factors to adjust self-reported estimates of obesity. *Public Health Reports, 19*(3), 71-82. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18847148

Greenberg, B. G., Kuebler, R. R., Jr., Abernathy, J. R., & Horvitz, D. G. (1971). Application of the Randomized Response Technique in obtaining quantitative data. *Journal of the American Statistical Association, 66*(334), 243-250. doi:10.2307/2283916

Griesler, P. C., Kandel, D. B., Schaffran, C., Hu, M. C., & Davies, M. (2008). Adolescents' inconsistency in self-reported smoking: A comparison of reports in school and in household settings. *Public Opinion Quarterly, 72*(2), 260-290. doi:10.1093/poq/nfn016

Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *Lancet, 359*(9302), 248-252. doi:10.1016/S0140-6736(02)07451-2

Hahn, S., Puffer, S., Torgerson, D. J., & Watson, J. (2005). Methodological bias in cluster randomised trials. *BMC Medical Research Methodology, 5*, 1-8. doi:10.1186/1471-2288-5-10

Hall, M. F. (1995). Patient satisfaction or acquiescence? Comparing mail and telephone survey results. *Journal of Health Care Marketing, 15*(1), 54-61. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10142388

Hebert, J. R., Ebbeling, C. B., Matthews, C. E., Hurley, T. G., MA, Y., Druker, S., & Clemow, L. (2002). Systematic errors in middle-aged women's estimates of energy intake: Comparing three self-report measures to total energy expenditure from doubly labeled water. *Annals of Epidemiology, 12*(8), 577-586. doi:10.1016/S1047-2797(01)00297-6

Heckman, J. T. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153-161. doi:10.2307/1912352

Hinz, A., Michalski, D., Schwarz, R., & Herzberg, P. Y. (2007). The acquiescence effect in responding to a questionnaire. *Psycho-Social-Medicine, 4*, 1-9. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19742288

Hsieh, S., Lee, S., Li, C., & Tu, S. (2016). An alternative to unrelated randomized response techniques with logistic regression analysis. *Statistical Methods & Applications, 25*(4), 601-621. Retrieved from https://link.springer.com/article/10.1007/s10260-016-0351-1

Hultsch, D. F., MacDonald, W. S., Hunter, M. A., Maitland, S. B., & Dixon, R. A. (2002). Sampling and generalisability in developmental research: Comparison of random and convenience samples of older adults. *International Journal of Behavioral Development, 26*(4), 345-359. doi:10.1080/01650250143000247

Iwata, N. (2014). Cultural distinctiveness in response bias. In M. Dollard, A. Shimazu, R. B. Nordin, P. Brough, & M. Tuckey (Eds.), *Psychosocial Factors at Work in the Asia Pacific*. Retrieved from https://link.springer.com/chapter/10.1007/978-94-017-8975-2_3.

Johnson, R. A., Gerstein, D. R., & Rasinski, K. A. (1998). Adjusting survey estimates for response bias: An application to trends in alcohol and marijuana use. *The Public Opinion Quarterly, 62*(3), 354-377. doi:10.1086/297850

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277. doi:10.1177/0022022104272905

Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry and Neuroscience, 37*(3), 149-152. doi:10.1503/jpn.120065

Juster, F. T., & Kuester, K. A. (1991). Differences in the measurement of wealth, wealth inequality and wealth composition obtained from alternative U.S. wealth surveys. *Review of Income and Wealth, 37*(1), 33-62. doi:10.1111/j.1475-4991.1991.tb00337.x

Keogh-Brown, M. R., Bachmann, M. O., Shepstone, L., Hewitt, C., Howe, A., Ramsay, C. R., . . . Campbell, M. J. (2007). Contamination in trials of educational interventions. *Health Technology Assessment, 11*(43), 1-107. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17935683

Krishna, R., Maithreyi, R., & Surpaneni, K. M. (2010). Research bias: A review for medical students. *Journal of Clinical and Diagnostic Research, 4*, 2320-2324. Retrieved from https://pdfs.semanticscholar.org/cda7/4cab25cfb58b0927ee926afc25997c4519a7.pdf

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47*(4), 2025-2047. doi:10.1007/s11135-011-9640-9

Lai, L. C. H., Cummins, R. A., & Lau, A. L. D. (2013). Cross-cultural differences in subjective wellbeing: Cultural response bias in explanation. *Social Indicators Research, 114*(2), 607-619. doi:10.1007/s11205-012-0164-z

Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The Marlowe-Crowne Social Desirability Scale outperforms BIDR Impression Management Scale for identifying fakers. *The Journal of Research in Personality, 61*, 80-86. doi:10.1016/j.jrp.2016.02.004

Langeland, W., Smit, J. H., Merckelbach, H., de Vries, G., Hoogendoorn, A. W., & Draijer, N. (2015). Inconsistent retrospective self-reports of childhood sexual abuse and their correlates in the general population. *Social Psychiatry and Psychiatric Epidemiology, 50*(4), 603-612. doi:10.1007/s00127-014-0986-x

Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity of scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement, 67*(5), 859-876. doi:10.1177/0013164406299104

Lauritsen, J. L., & Swicegood, C. G. (1997). The consistency of self-reported initiation of sexual activity. *Family Planning Perspectives, 29*(5), 215-221. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9323498

Lavrakas, P. J. (2008a). Acquiescence bias. *Encyclopedia of Survey Research Methods*. Retrieved from http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n3.xml

Lavrakas, P. J. (2008b). Response bias. *Encyclopedia of Survey Research Methods*. Retrieved from http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n486.xml

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., . . . Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine, 367*(14), 1355-1360. doi:10.1056/NEJMsr1203730

Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Applied Social Psychology, 34*(11), 2343-2352. doi:10.1111/j.1559-1816.2004.tb01980.x

Mazor, K. M., Clauser, B. E., Field, T., Yood, R. A., & Gurwitz, J. H. (2002). A demonstration of the impact of response bias on the results of patient satisfaction surveys. *Health Services Research Journal, 37*(5), 1403-1417. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12479503

Melnikow, J., & Kiefe, C. (1994). Patient compliance and medical research: Issues in methodology. *Journal of General Internal Medicine, 9*(2), 96-105. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8164085

Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M., & Cook, N. E. (2013). Careless and random responding on clinical and research measures in the addictions: A concerning problem and investigation of their detection. *International Journal of Mental Health and Addiction, 11*(3), 292-306. doi:10.1007/s11469-012-9410-5

Myers, W. R. (2000). Handling missing data in clinical trials: An overview. *Drug Information Journal, 34*(2), 525-533. doi:10.1177/009286150003400221

Navarro-González, D., Lorenzo-Seva, U., & Vigil-Colet, A. (2016). How response bias affects the factorial structure of personality self-reports. *Psicothema, 28*(4), 465-470. doi:10.7334/psicothema2016.113

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263-280. doi:10.1002/ejsp.2420150303

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology, 135*(2), 151-165. doi:10.3200/GENP.135.2.151-166

O'Leary, T. E., Diller, L., & Recklitis, C. J. (2007). The effects of response bias on self-reported quality of life among childhood cancer survivors. *Quality of Life Research, 16*(7), 1211-1220. doi:10.1007/s11136-007-9231-3

Ong, A. D., & Weiss, D. J. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology, 30*(8), 1691-1708. doi:10.1111/j.1559-1816.2000.tb02462.x

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776-783. doi:10.1037/h0043424

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and avoiding bias in research. *Plastic and Reconstructive Surgery, 126*(2), 619-625. doi:10.1097/PRS.0b013e3181de24bc

Paulhus, D. L. (1998). *Manual for the Balanced Inventory of Desirable Responding (BIDR-7)*. Toronto, Ontario, Canada: Multi-Health Systems.

Pfeffer, F. T., & Griffin, J. (2017). Determinants of wealth fluctuation: Changes in hard-to-measure economic variables in a panel study. *Methods, Data, Anayzes, 11*(1), 87-108. doi:10.12758/mda.2016.015

Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from Bed to Bench, 5*(2), 79-83. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24834204

Preston, S. H., Fishman, E., & Stokes, A. (2015). Effects of categorization and self-report bias on estimates of the association between obesity and mortality. *Annals of Epidemiology, 25*(12), 907-911.e901-902. doi:10.1016/j.annepidem.2015.07.012

Pruchno, R. A., Brill, J. E., Shands, Y., Gordon, J. R., Genderson, M. W., Rose, M., & Cartwright, F. (2008). Convenience samples and caregiving research: How generalizable are the findings? *The Gerontologist, 48*(6), 820-827. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19139255

Riphahn, R. T., & Serfling, O. (2005). Item non-response on income and wealth questions. *Empirical Economics, 30*(2), 521-538. doi:10.1007/s00181-005-0247-7

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*(2), 363-375. doi:10.1037/0033-2909.114.2.363

Sackett, D. L. (2007). Commentary: Measuring the success of blinding in RCTs: Don't, must, can't or needn't? *International Journal of Epidemiology, 36*(3), 664-665. doi:10.1093/ije/dym088

Sackett, D. L. (2011). Clinician-trialist rounds: 5. Cointervention bias--how to diagnose it in their trial and prevent it in yours. *Clinical Trials, 8*(4), 440-442. doi:10.1177/1740774511410995

Saczynski, J. S., McManus, D. D., & Goldberg, R. J. (2013). Commonly utilized data collection approaches in clinical research. *The American Journal of Medicine, 126*(11), 10. doi:10.1016/j.amjmed.2013.04.016

Schulz, K. F., & Grimes, D. A. (2002). Blinding in randomised trials: Hiding who got what. *The Lancet, 359*(9307), 696-700. doi:10.1016/S0140-6736(02)07816-9

Shields, M., Gorber, S. C., & Tremblay, M. S. (2008). Estimates of obesity based on self-report versus direct measures. *Public Health Reports, 19*(2), 61-76. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18642520

Siegel, D. M., Aten, M. J., & Roghmann, K. J. (1998). Self-reported honesty among middle and high school students responding to a sexual behavior questionnaire. *Journal of Adolescent Health, 23*(1), 20-28. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9648019

Skelly, A. C., Dettori, J. R., & Brodt, E. D. (2012). Assessing bias: The importance of considering confounding. *Evidence-Based Spine-Care Journal, 3*(1), 9-12. doi:10.1055/s-0031-1298595

Smith, T. W. (1991). Adult sexual behavior in 1989: Number of partners, frequency of intercourse and risk of AIDS. *Family Planning Perspectives, 23*(3), 102-107. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1860475

Szklo, M., & Nieto, F. J. (2014). Understanding lack of validity: Bias. In M. Johnson & T. Reilly (Eds.), *Epidemiology: Beyond the Basics* (pp. 109-151). Burlington, MA: Jones & Bartlett Learning.

Toneatto, T., Sobell, L. C., & Sobell, M. B. (1992). Predictors of alcohol abusers' inconsistent self-reports of their drinking and life events. *Alcoholism: Clinical and Experimental Research, 16*(3), 542-546. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1626653

Torgerson, D. J. (2001). Contamination in trials: Is cluster randomisation the answer? *The BMJ, 322*(7282), 355-357. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11159665

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859-883. doi:10.1037/0033-2909.133.5.859

Tripepi, G., Jager, K. J., Dekker, F. W., Wanner, C., & Zoccali, C. (2008). Bias in clinical research. *Kidney International, 73*(2), 148-153. doi:10.1038/sj.ki.5002648

Turrell, G. (2000). Income non-reporting: Implications for health inequalities research. *Journal of Epidemiology and Community Health, 54*(3), 207-214. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10746115 van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.818.5855&rep=rep1&type=pdf

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). Estimation. In C. Crockett, B. Gershman, & C. Ronquillo (Eds.), *Mathematical statistics with applications* (pp. 390-443). Belmont, CA: Brooks/Cole.

Walsh-Buhi, E. R., Marhefka, S. L., Wang, W., Debate, R., Perrin, K., Singleton, A., . . . Daley, E. M. (2016). The impact of the Teen Outreach Program on sexual intentions and behaviors. *Journal of Adolescent Health, 59*(3), 283-290. doi:10.1016/j.jadohealth.2016.05.007

Ward, Z. J., Long, M. W., Resch, S. C., Gortmaker, S. L., Cradock, A. L., Giles, C., . . . Wang, Y. C. (2016). Redrawing the US obesity landscape: Bias-corrected estimates of state-specific adult obesity prevalence. *PLoS One, 11*(3), e0150735. doi:10.1371/journal.pone.0150735

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63-66. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12261830

Wiederman, M. W. (1997). The truth must be in here somewhere: Examining the gender discrepancy in self-reported lifetime number of sex partners. *The Journal of Sex Research, 34*(4), 375-386. Retrieved from http://www.jstor.org/stable/3813479?seq=1#page_scan_tab_contents

Wouters, K., Maesschalck, J., Peeters, C. F. W., & Roosen, M. (2014). Methodological issues in the design of online surveys for measuring unethical work behavior: Recommendations on the basis of a split-ballot experiment. *Journal of Business Ethics, 120*(2), 275-289. doi:10.1007/s10551-013-1659-5

Yu, Q., Stasny, E. A., & Li, B. (2008). Bayesian models to adjust for response bias in survey data for estimating rape and domestic violence rates from the NCVS. *The Annals of Applied Statistics, 2*(2), 665-686. doi:10.1007/s10551-013-1659-5

Zhang, X., Kuchinke, L., Woud, M. L., Velten, J., & Margraf, J. (2017). Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior, 71*, 172-180. doi:10.1016/j.chb.2017.02.006

Zimmerman, R. S., & Langer, L. M. (1995). Improving estimates of prevalence rates of sensitive behaviors: The Randomized Lists Technique and consideration of self-reported honesty. *The Journal of Sex Research, 32*(2), 107-117. doi:10.1080/00224499509551781

Appendices

The R code used for data generation, bias correction models, bias and *RRMSE* (%), and *RRMSE* (%) figures are presented below in Appendices A, B, C, and D, respectively. All codes (excluding that for figures) were rerun multiple times with changes to sample size, effect size, and sample reliability – all of which are highlighted in yellow to show where the changes were made – in order to generate the estimates for the various scenarios. The code for sample size appears on page 70 in the "Sample" section of the code, currently as "n <- 20000" for the 20000 sample size, but was also run as "n <- 2000" and "n <- 200" for sample sizes of 2000 and 200, respectively. The code for sample reliability appears on page 70 in the "Reliability Distribution" section of the code, currently as "rlb.star <- abs(rnorm(n, 2, 1))" for a sample reliability of 80%, but was also run as "rlb.star <- abs(rnorm(n, 1, 1))" for a sample reliability of 50%. As this code was updated, the histograms generated for Figure 2 were also updated. The code for effect size appears in two locations, the first on page 70 in the "Parameters" section of the code and the second on page 89 in the "Bias and *RRMSE* (%)" section. These codes are currently written as "b1 <- -1.4204" and "or <- -1.4204," respectively for the large effect size, but were also run as "b1 <- -0.9152" and "or <- -0.9152" for the medium effect size and "b1 <- -0.3795" and "or <- -0.3795" for the small effect size. As a reminder, these are the log odds of the inverse *ORs* that correspond to a baseline prevalence rate of 10%. Due to the length of the code, some sections were formatted into columns to save space.

Appendix A: R Code for Data Generation

```
library(xlsx)
library(lattice)
library(ggplot2)

set.seed(1234)

rep <- 1000


######################### Sample #########################

n <- 20000
n1 <- n2 <- n / 2
tx <- c ( rep ( 0, n1 ), rep ( 1, n2 ) )

prop.a <- .5
prop.b <- 1 - prop.a
group <- rbinom ( n, 1, prop.a )


################## Reliability Distribution ###################

est2 <- matrix (0, rep, 265)

for ( i in ( 1 : rep ) ) {

r.rlb <- .75 # ratio of reliabilities between the two groups; group a vs group b
rlb.star <- abs ( rnorm ( n, 2, 1 ) )
rlb.star [ group == 1 ] <- rlb.star [ group == 1 ] * r.rlb
rlb <- rlb.star; rlb [rlb > 1 ] <- 1

## histograms ## - FIGURE 2

histogram (rlb, main="Reliability Distribution \n N=20000 R~80%",
xlab="Probability of Responding Honestly", ylab="Percentage of Total Sample",
ylim=c(0,100), breaks=c(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0), col=c("azure3"))

histogram (rlb [rlb<1], main="Unreliability Distribution \n N=20000 R~80%",
xlab="Probability of Responding Honestly", ylab="Percentage of Unreliable Sample",
ylim=c(0,25), breaks=c(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0), col=c("azure3"))

summary(rlb)
summary(rlb==1)

# rnorm (n, 1, 1) --> rel ~ 50%
# rnorm (n, 2, 1) --> rel ~ 80%
# Group 0 = Group A
# Group 1 = Group B
# Group B is approx. 25% less reliable than group A

# convert reliability measure to a binomial decision
rlb.b <- rbinom (n , 1, rlb )


####################### Parameters #######################

# simulate regression coefficients

b0  <-  -1.0          # exp(-1.0) = odds (0.3679) = probability (0.27)
b1  <-  -1.4204       # cohen's d - see below
b2  <-   0.75         # exp(0.75) = 2.12
```

```
# values for b1 corresponding to OR strengths / cohen's d:
	# small:		b1 = -0.3795 --> exp(-0.3795) = OR = 0.6842 (inverse OR = 1.4615)
	# medium:		b1 = -0.9152 --> exp(-0.9152) = OR = 0.4004 (inverse OR = 2.4972)
	# large:		b1 = -1.4204 --> exp(-1.4204) = OR = 0.2416 (inverse OR = 4.1387)
# ORs based on event rate = 10% in non-exposed
# See Chen, Cohen, & Chen (2010) for details

predictor <- group
table ( predictor )

# simulate outcome that incorporates tx and predictor

z <- b0 + b1*tx + b2*predictor				#linear combination
invlogit <- ( exp(z) / ( 1 + exp(z)))			#probability of outcome

yy <- rbinom(n,1,invlogit)

table(yy)


################ Unreliable Response Styles #################


###### Predictor

# careless responding (arbitrary)

p.arb <- .5
predict.arb <- rbinom (n , 1, p.arb )

# opposite
predict.ops <- 1 - predictor

# social desirability
predict.social <- rep (0, n )

# incorporate rlb to predictor response

predictor1 <- predictor
predictor1 [ rlb.b == 0 ] <- predict.arb [ rlb.b == 0 ]
# verify
cbind ( predictor, predict.arb, rlb.b, predictor1 ) [ 1 : 100, ]

predictor2 <- predictor
predictor2 [ rlb.b == 0 ] <- predict.ops [ rlb.b == 0 ]

predictor3 <- predictor
predictor3 [ rlb.b == 0 ] <- predict.social [ rlb.b == 0 ]

table(predictor1)
table(predictor2)
table(predictor3)
```

```
###### Outcome

# careless responding (arbitrary)

p.arb <- .5
yy.arb <- rbinom (n , 1, p.arb )

# opposite
yy.ops <- 1 - yy

# social desirability
yy.social <- rep (0, n )

# incorporate rlb to yy response

yy1 <- yy
yy1 [ rlb.b == 0 ] <- yy.arb [ rlb.b == 0 ]
# verify
cbind ( yy, yy.arb, rlb.b, yy1 ) [ 1 : 100, ]

yy2 <- yy
yy2 [ rlb.b == 0 ] <- yy.ops [ rlb.b == 0 ]

yy3 <- yy
yy3 [ rlb.b == 0 ] <- yy.social [ rlb.b == 0 ]

table(yy1)
table(yy2)
table(yy3)
```

Appendix B: R Code for Bias Correction Models

```
#################### Unadjusted #####################

# true y
R1 <-  glm ( yy ~ tx + predictor  , fam = binomial )
summary(R1)
est2 [i, 1] <-  exp ( R1 $ coefficients [ "tx" ] )

# arb y
R2 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial )
summary(R2)
est2 [i, 2] <-  exp ( R2 $ coefficients [ "tx" ] )

# ops y
R3 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial )
summary(R3)
est2 [i, 3] <-  exp ( R3 $ coefficients [ "tx" ] )

# social y
R4 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial )
summary(R4)
est2 [i, 4] <-  exp ( R4 $ coefficients [ "tx" ] )


##################### Threshold Only ###################

### Cut at rlb = 1
cut <- 1

# arb y
R5 <- glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut ))
summary(R5)
est2 [i, 5] <-  exp ( R5 $ coefficients [ "tx" ] )

# ops y
R6 <- glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut ))
summary(R6)
est2 [i, 6] <-  exp ( R6 $ coefficients [ "tx" ] )

# social y
R7 <- glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut ))
summary(R7)
est2 [i, 7] <-  exp ( R7 $ coefficients [ "tx" ] )

### Cut at rlb = .95
cut <- .95

# arb y
R11 <- glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R11)
est2 [i, 11] <-  exp ( R11 $ coefficients [ "tx" ] )

# ops y
R12 <- glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R12)
est2 [i, 12] <-  exp ( R12 $ coefficients [ "tx" ] )

# social y
R13 <- glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R13)
est2 [i, 13] <-  exp ( R13 $ coefficients [ "tx" ] )


### Cut at rlb = .99
cut <- .99

# arb y
R8 <- glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut ))
summary(R8)
est2 [i, 8] <-  exp ( R8 $ coefficients [ "tx" ] )

# ops y
R9 <- glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut ))
summary(R9)
est2 [i, 9] <-  exp ( R9 $ coefficients [ "tx" ] )

# social y
R10 <- glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R10)
est2 [i, 10] <-  exp ( R10 $ coefficients [ "tx" ] )

### Cut at rlb = .9
cut <- .9

# arb y
R14 <- glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R14)
est2 [i, 14] <-  exp ( R14 $ coefficients [ "tx" ] )

# ops y
R15 <- glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R15)
est2 [i, 15] <-  exp ( R15 $ coefficients [ "tx" ] )

# social y
R16 <- glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R16)
est2 [i, 16] <-  exp ( R16 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .85
cut <- .85

# arb y
R17 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R17)
est2 [i, 17] <-  exp ( R17 $ coefficients [ "tx" ] )

# ops y
R18 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R18)
est2 [i, 18] <-  exp ( R18 $ coefficients [ "tx" ] )

# social y
R19 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R19)
est2 [i, 19] <-  exp ( R19 $ coefficients [ "tx" ] )


### Cut at rlb = .8
cut <- .8

# arb y
R20 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R20)
est2 [i, 20] <-  exp ( R20 $ coefficients [ "tx" ] )

# ops y
R21 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R21)
est2 [i, 21] <-  exp ( R21 $ coefficients [ "tx" ] )

# social y
R22 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R22)
est2 [i, 22] <-  exp ( R22 $ coefficients [ "tx" ] )


### Cut at rlb = .75
cut <- .75

# arb y
R23 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R23)
est2 [i, 23] <-  exp ( R23 $ coefficients [ "tx" ] )

# ops y
R24 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R24)
est2 [i, 24] <-  exp ( R24 $ coefficients [ "tx" ] )

# social y
R25 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R25)
est2 [i, 25] <-  exp ( R25 $ coefficients [ "tx" ] )


### Cut at rlb = .7
cut <- .7

# arb y
R26 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R26)
est2 [i, 26] <-  exp ( R26 $ coefficients [ "tx" ] )

# ops y
R27 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R27)
est2 [i, 27] <-  exp ( R27 $ coefficients [ "tx" ] )

# social y
R28 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R28)
est2 [i, 28] <-  exp ( R28 $ coefficients [ "tx" ] )


### Cut at rlb = .65
cut <- .65

# arb y
R29 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R29)
est2 [i, 29] <-  exp ( R29 $ coefficients [ "tx" ] )

# ops y
R30 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R30)
est2 [i, 30] <-  exp ( R30 $ coefficients [ "tx" ] )

# social y
R31 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R31)
est2 [i, 31] <-  exp ( R31 $ coefficients [ "tx" ] )


### Cut at rlb = .6
cut <- .6

# arb y
R32 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R32)
est2 [i, 32] <-  exp ( R32 $ coefficients [ "tx" ] )

# ops y
R33 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R33)
est2 [i, 33] <-  exp ( R33 $ coefficients [ "tx" ] )

# social y
R34 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R34)
est2 [i, 34] <-  exp ( R34 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .55
cut <- .55

# arb y
R35 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R35)
est2 [i, 35] <-  exp ( R35 $ coefficients [ "tx" ] )

# ops y
R36 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R36)
est2 [i, 36] <-  exp ( R36 $ coefficients [ "tx" ] )

# social y
R37 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R37)
est2 [i, 37] <-  exp ( R37 $ coefficients [ "tx" ] )



### Cut at rlb = .5
cut <- .5

# arb y
R38 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R38)
est2 [i, 38] <-  exp ( R38 $ coefficients [ "tx" ] )

# ops y
R39 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R39)
est2 [i, 39] <-  exp ( R39 $ coefficients [ "tx" ] )

# social y
R40 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R40)
est2 [i, 40] <-  exp ( R40 $ coefficients [ "tx" ] )



### Cut at rlb = .45
cut <- .45

# arb y
R41 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R41)
est2 [i, 41] <-  exp ( R41 $ coefficients [ "tx" ] )

# ops y
R42 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R42)
est2 [i, 42] <-  exp ( R42 $ coefficients [ "tx" ] )

# social y
R43 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R43)
est2 [i, 43] <-  exp ( R43 $ coefficients [ "tx" ] )

### Cut at rlb = .4
cut <- .4

# arb y
R44 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R44)
est2 [i, 44] <-  exp ( R44 $ coefficients [ "tx" ] )

# ops y
R45 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R45)
est2 [i, 45] <-  exp ( R45 $ coefficients [ "tx" ] )

# social y
R46 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R46)
est2 [i, 46] <-  exp ( R46 $ coefficients [ "tx" ] )



### Cut at rlb = .35
cut <- .35

# arb y
R47 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R47)
est2 [i, 47] <-  exp ( R47 $ coefficients [ "tx" ] )

# ops y
R48 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R48)
est2 [i, 48] <-  exp ( R48 $ coefficients [ "tx" ] )

# social y
R49 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R49)
est2 [i, 49] <-  exp ( R49 $ coefficients [ "tx" ] )



### Cut at rlb = .3
cut <- .3

# arb y
R50 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R50)
est2 [i, 50] <-  exp ( R50 $ coefficients [ "tx" ] )

# ops y
R51 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R51)
est2 [i, 51] <-  exp ( R51 $ coefficients [ "tx" ] )

# social y
R52 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R52)
est2 [i, 52] <-  exp ( R52 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .25
cut <- .25

# arb y
R53 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R53)
est2 [i, 53] <-  exp ( R53 $ coefficients [ "tx" ] )

# ops y
R54 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R54)
est2 [i, 54] <-  exp ( R54 $ coefficients [ "tx" ] )

# social y
R55 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R55)
est2 [i, 55] <-  exp ( R55 $ coefficients [ "tx" ] )



### Cut at rlb = .2
cut <- .2

# arb y
R56 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R56)
est2 [i, 56] <-  exp ( R56 $ coefficients [ "tx" ] )

# ops y
R57 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R57)
est2 [i, 57] <-  exp ( R57 $ coefficients [ "tx" ] )

# social y
R58 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R58)
est2 [i, 58] <-  exp ( R58 $ coefficients [ "tx" ] )



### Cut at rlb = .15
cut <- .15

# arb y
R59 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R59)
est2 [i, 59] <-  exp ( R59 $ coefficients [ "tx" ] )

# ops y
R60 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R60)
est2 [i, 60] <-  exp ( R60 $ coefficients [ "tx" ] )

# social y
R61 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R61)
est2 [i, 61] <-  exp ( R61 $ coefficients [ "tx" ] )

### Cut at rlb = .1
cut <- .1

# arb y
R62 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R62)
est2 [i, 62] <-  exp ( R62 $ coefficients [ "tx" ] )

# ops y
R63 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R63)
est2 [i, 63] <-  exp ( R63 $ coefficients [ "tx" ] )

# social y
R64 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R64)
est2 [i, 64] <-  exp ( R64 $ coefficients [ "tx" ] )



### Cut at rlb = .05
cut <- .05

# arb y
R65 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , sub = ( rlb >= cut))
summary(R65)
est2 [i, 65] <-  exp ( R65 $ coefficients [ "tx" ] )

# ops y
R66 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , sub = ( rlb >= cut))
summary(R66)
est2 [i, 66] <-  exp ( R66 $ coefficients [ "tx" ] )

# social y
R67 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , sub = ( rlb >= cut))
summary(R67)
est2 [i, 67] <-  exp ( R67 $ coefficients [ "tx" ] )
```

```
###################### Weight Only #####################


### Linear Weight = rlb

# arb y
R68 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb )
summary(R68)
est2 [i, 68] <-  exp ( R68 $ coefficients [ "tx" ] )

# ops y
R69 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb )
summary(R69)
est2 [i, 69] <-  exp ( R69 $ coefficients [ "tx" ] )

# social y
R70 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb )
summary(R70)
est2 [i, 70] <-  exp ( R70 $ coefficients [ "tx" ] )



### Quadratic Weight = rlb ^2

# arb y
R71 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2 )
summary(R71)
est2 [i, 71] <-  exp ( R71 $ coefficients [ "tx" ] )

# ops y
R72 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2 )
summary(R72)
est2 [i, 72] <-  exp ( R72 $ coefficients [ "tx" ] )

# social y
R73 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2 )
summary(R73)
est2 [i, 73] <-  exp ( R73 $ coefficients [ "tx" ] )



### Cubic Weight = rlb ^3

# arb y
R74 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3 )
summary(R74)
est2 [i, 74] <-  exp ( R74 $ coefficients [ "tx" ] )

# ops y
R75 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3 )
summary(R75)
est2 [i, 75] <-  exp ( R75 $ coefficients [ "tx" ] )

# social y
R76 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3 )
summary(R76)
est2 [i, 76] <-  exp ( R76 $ coefficients [ "tx" ] )
```

```
################# Threshold & Linear Weight ################

### Cut at rlb = 1 and linear weight
cut <- 1

# arb y
R77 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R77)
est2 [i, 77] <-  exp ( R77 $ coefficients [ "tx" ] )

# ops y
R78 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R78)
est2 [i, 78] <-  exp ( R78 $ coefficients [ "tx" ] )

# social y
R79 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R79)
est2 [i, 79] <-  exp ( R79 $ coefficients [ "tx" ] )


### Cut at rlb = .99 and linear weight
cut <- .99

# arb y
R80 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R80)
est2 [i, 80] <-  exp ( R80 $ coefficients [ "tx" ] )

# ops y
R81 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R81)
est2 [i, 81] <-  exp ( R81 $ coefficients [ "tx" ] )

# social y
R82 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R82)
est2 [i, 82] <-  exp ( R82 $ coefficients [ "tx" ] )


### Cut at rlb = .95 and linear weight
cut <- .95

# arb y
R83 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R83)
est2 [i, 83] <-  exp ( R83 $ coefficients [ "tx" ] )

# ops y
R84 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R84)
est2 [i, 84] <-  exp ( R84 $ coefficients [ "tx" ] )

# social y
R85 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R85)
est2 [i, 85] <-  exp ( R85 $ coefficients [ "tx" ] )

### Cut at rlb = .9 and linear weight
cut <- .9

# arb y
R86 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R86)
est2 [i, 86] <-  exp ( R86 $ coefficients [ "tx" ] )

# ops y
R87 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R87)
est2 [i, 87] <-  exp ( R87 $ coefficients [ "tx" ] )

# social y
R88 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R88)
est2 [i, 88] <-  exp ( R88 $ coefficients [ "tx" ] )


### Cut at rlb = .85 and linear weight
cut <- .85

# arb y
R89 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R89)
est2 [i, 89] <-  exp ( R89 $ coefficients [ "tx" ] )

# ops y
R90 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R90)
est2 [i, 90] <-  exp ( R90 $ coefficients [ "tx" ] )

# social y
R91 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R91)
est2 [i, 91] <-  exp ( R91 $ coefficients [ "tx" ] )


### Cut at rlb = .8 and linear weight
cut <- .8

# arb y
R92 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R92)
est2 [i, 92] <-  exp ( R92 $ coefficients [ "tx" ] )

# ops y
R93 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R93)
est2 [i, 93] <-  exp ( R93 $ coefficients [ "tx" ] )

# social y
R94 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R94)
est2 [i, 94] <-  exp ( R94 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .75 and linear weight
cut <- .75

# arb y
R95 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R95)
est2 [i, 95] <-  exp ( R95 $ coefficients [ "tx" ] )

# ops y
R96 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R96)
est2 [i, 96] <-  exp ( R96 $ coefficients [ "tx" ] )

# social y
R97 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R97)
est2 [i, 97] <-  exp ( R97 $ coefficients [ "tx" ] )


### Cut at rlb = .7 and linear weight
cut <- .7

# arb y
R98 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R98)
est2 [i, 98] <-  exp ( R98 $ coefficients [ "tx" ] )

# ops y
R99 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R99)
est2 [i, 99] <-  exp ( R99 $ coefficients [ "tx" ] )

# social y
R100 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R100)
est2 [i, 100] <-  exp ( R100 $ coefficients [ "tx" ] )


### Cut at rlb = .65 and linear weight
cut <- .65

# arb y
R101 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R101)
est2 [i, 101] <-  exp ( R101 $ coefficients [ "tx" ] )

# ops y
R102 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R102)
est2 [i, 102] <-  exp ( R102 $ coefficients [ "tx" ] )

# social y
R103 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R103)
est2 [i, 103] <-  exp ( R103 $ coefficients [ "tx" ] )

### Cut at rlb = .6 and linear weight
cut <- .6

# arb y
R104 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R104)
est2 [i, 104] <-  exp ( R104 $ coefficients [ "tx" ] )

# ops y
R105 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R105)
est2 [i, 105] <-  exp ( R105 $ coefficients [ "tx" ] )

# social y
R106 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R106)
est2 [i, 106] <-  exp ( R106 $ coefficients [ "tx" ] )


### Cut at rlb = .55 and linear weight
cut <- .55

# arb y
R107 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R107)
est2 [i, 107] <-  exp ( R107 $ coefficients [ "tx" ] )

# ops y
R108 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R108)
est2 [i, 108] <-  exp ( R108 $ coefficients [ "tx" ] )

# social y
R109 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R109)
est2 [i, 109] <-  exp ( R109 $ coefficients [ "tx" ] )


### Cut at rlb = .5 and linear weight
cut <- .5

# arb y
R110 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R110)
est2 [i, 110] <-  exp ( R110 $ coefficients [ "tx" ] )

# ops y
R111 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R111)
est2 [i, 111] <-  exp ( R111 $ coefficients [ "tx" ] )

# social y
R112 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R112)
est2 [i, 112] <-  exp ( R112 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .45 and linear weight
cut <- .45

# arb y
R113 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R113)
est2 [i, 113] <-  exp ( R113 $ coefficients [ "tx" ] )

# ops y
R114 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R114)
est2 [i, 114] <-  exp ( R114 $ coefficients [ "tx" ] )

# social y
R115 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R115)
est2 [i, 115] <-  exp ( R115 $ coefficients [ "tx" ] )


### Cut at rlb = .4 and linear weight
cut <- .4

# arb y
R116 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R116)
est2 [i, 116] <-  exp ( R116 $ coefficients [ "tx" ] )

# ops y
R117 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R117)
est2 [i, 117] <-  exp ( R117 $ coefficients [ "tx" ] )

# social y
R118 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R118)
est2 [i, 118] <-  exp ( R118 $ coefficients [ "tx" ] )


### Cut at rlb = .35 and linear weight
cut <- .35

# arb y
R119 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R119)
est2 [i, 119] <-  exp ( R119 $ coefficients [ "tx" ] )

# ops y
R120 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R120)
est2 [i, 120] <-  exp ( R120 $ coefficients [ "tx" ] )

# social y
R121 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R121)
est2 [i, 121] <-  exp ( R121 $ coefficients [ "tx" ] )

### Cut at rlb = .3 and linear weight
cut <- .3

# arb y
R122 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R122)
est2 [i, 122] <-  exp ( R122 $ coefficients [ "tx" ] )

# ops y
R123 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R123)
est2 [i, 123] <-  exp ( R123 $ coefficients [ "tx" ] )

# social y
R124 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
 sub = ( rlb >= cut ))
summary(R124)
est2 [i, 124] <-  exp ( R124 $ coefficients [ "tx" ] )


### Cut at rlb = .25 and linear weight
cut <- .25

# arb y
R125 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R125)
est2 [i, 125] <-  exp ( R125 $ coefficients [ "tx" ] )

# ops y
R126 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R126)
est2 [i, 126] <-  exp ( R126 $ coefficients [ "tx" ] )

# social y
R127 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R127)
est2 [i, 127] <-  exp ( R127 $ coefficients [ "tx" ] )


### Cut at rlb = .2 and linear weight
cut <- .2

# arb y
R128 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R128)
est2 [i, 128] <-  exp ( R128 $ coefficients [ "tx" ] )

# ops y
R129 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R129)
est2 [i, 129] <-  exp ( R129 $ coefficients [ "tx" ] )

# social y
R130 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R130)
est2 [i, 130] <-  exp ( R130 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .15 and linear weight
cut <- .15

# arb y
R131 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R131)
est2 [i, 131] <-  exp ( R131 $ coefficients [ "tx" ] )

# ops y
R132 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R132)
est2 [i, 132] <-  exp ( R132 $ coefficients [ "tx" ] )

# social y
R133 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R133)
est2 [i, 133] <-  exp ( R133 $ coefficients [ "tx" ] )



### Cut at rlb = .1 and linear weight
cut <- .1

# arb y
R134 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R134)
est2 [i, 134] <-  exp ( R134 $ coefficients [ "tx" ] )

# ops y
R135 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R135)
est2 [i, 135] <-  exp ( R135 $ coefficients [ "tx" ] )

# social y
R136 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
 sub = ( rlb >= cut ))
summary(R136)
est2 [i, 136] <-  exp ( R136 $ coefficients [ "tx" ] )



### Cut at rlb = .05 and linear weight
cut <- .05

# arb y
R137 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R137)
est2 [i, 137] <-  exp ( R137 $ coefficients [ "tx" ] )

# ops y
R138 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R138)
est2 [i, 138] <-  exp ( R138 $ coefficients [ "tx" ] )

# social y
R139 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb,
sub = ( rlb >= cut ))
summary(R139)
est2 [i, 139] <-  exp ( R139 $ coefficients [ "tx" ] )
```

```
############### Threshold & Quadratic Weight ###############

### Cut at rlb = 1 and quadratic weight
cut <- 1

# arb y
R140 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R140)
est2 [i, 140] <-  exp ( R140 $ coefficients [ "tx" ] )

# ops y
R141 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R141)
est2 [i, 141] <-  exp ( R141 $ coefficients [ "tx" ] )

# social y
R142 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R142)
est2 [i, 142] <-  exp ( R142 $ coefficients [ "tx" ] )


### Cut at rlb = .99 and quadratic weight
cut <- .99

# arb y
R143 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R143)
est2 [i, 143] <-  exp ( R143 $ coefficients [ "tx" ] )

# ops y
R144 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R144)
est2 [i, 144] <-  exp ( R144 $ coefficients [ "tx" ] )

# social y
R145 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R145)
est2 [i, 145] <-  exp ( R145 $ coefficients [ "tx" ] )


### Cut at rlb = .95 and quadratic weight
cut <- .95

# arb y
R146 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R146)
est2 [i, 146] <-  exp ( R146 $ coefficients [ "tx" ] )

# ops y
R147 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R147)
est2 [i, 147] <-  exp ( R147 $ coefficients [ "tx" ] )

# social y
R148 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R148)
est2 [i, 148] <-  exp ( R148 $ coefficients [ "tx" ] )

### Cut at rlb = .9 and quadratic weight
cut <- .9

# arb y
R149 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R149)
est2 [i, 149] <-  exp ( R149 $ coefficients [ "tx" ] )

# ops y
R150 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R150)
est2 [i, 150] <-  exp ( R150 $ coefficients [ "tx" ] )

# social y
R151 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R151)
est2 [i, 151] <-  exp ( R151 $ coefficients [ "tx" ] )


### Cut at rlb = .85 and quadratic weight
cut <- .85

# arb y
R152 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R152)
est2 [i, 152] <-  exp ( R152 $ coefficients [ "tx" ] )

# ops y
R153 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R153)
est2 [i, 153] <-  exp ( R153 $ coefficients [ "tx" ] )

# social y
R154 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R154)
est2 [i, 154] <-  exp ( R154 $ coefficients [ "tx" ] )


### Cut at rlb = .8 and quadratic weight
cut <- .8

# arb y
R155 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R155)
est2 [i, 155] <-  exp ( R155 $ coefficients [ "tx" ] )

# ops y
R156 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R156)
est2 [i, 156] <-  exp ( R156 $ coefficients [ "tx" ] )

# social y
R157 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R157)
est2 [i, 157] <-  exp ( R157 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .75 and quadratic weight
cut <- .75

# arb y
R158 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R158)
est2 [i, 158] <-  exp ( R158 $ coefficients [ "tx" ] )

# ops y
R159 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R159)
est2 [i, 159] <-  exp ( R159 $ coefficients [ "tx" ] )

# social y
R160 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R160)
est2 [i, 160] <-  exp ( R160 $ coefficients [ "tx" ] )


### Cut at rlb = .7 and quadratic weight
cut <- .7

# arb y
R161 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R161)
est2 [i, 161] <-  exp ( R161 $ coefficients [ "tx" ] )

# ops y
R162 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R162)
est2 [i, 162] <-  exp ( R162 $ coefficients [ "tx" ] )

# social y
R163 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R163)
est2 [i, 163] <-  exp ( R163 $ coefficients [ "tx" ] )


### Cut at rlb = .65 and quadratic weight
cut <- .65

# arb y
R164 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R164)
est2 [i, 164] <-  exp ( R164 $ coefficients [ "tx" ] )

# ops y
R165 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R165)
est2 [i, 165] <-  exp ( R165 $ coefficients [ "tx" ] )

# social y
R166 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R166)
est2 [i, 166] <-  exp ( R166 $ coefficients [ "tx" ] )

### Cut at rlb = .6 and quadratic weight
cut <- .6

# arb y
R167 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R167)
est2 [i, 167] <-  exp ( R167 $ coefficients [ "tx" ] )

# ops y
R168 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R168)
est2 [i, 168] <-  exp ( R168 $ coefficients [ "tx" ] )

# social y
R169 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R169)
est2 [i, 169] <-  exp ( R169 $ coefficients [ "tx" ] )


### Cut at rlb = .55 and quadratic weight
cut <- .55

# arb y
R170 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R170)
est2 [i, 170] <-  exp ( R170 $ coefficients [ "tx" ] )

# ops y
R171 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R171)
est2 [i, 171] <-  exp ( R171 $ coefficients [ "tx" ] )

# social y
R172 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R172)
est2 [i, 172] <-  exp ( R172 $ coefficients [ "tx" ] )


### Cut at rlb = .5 and quadratic weight
cut <- .5

# arb y
R173 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R173)
est2 [i, 173] <-  exp ( R173 $ coefficients [ "tx" ] )

# ops y
R174 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R174)
est2 [i, 174] <-  exp ( R174 $ coefficients [ "tx" ] )

# social y
R175 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R175)
est2 [i, 175] <-  exp ( R175 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .45 and quadratic weight
cut <- .45

# arb y
R176 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R176)
est2 [i, 176] <-  exp ( R176 $ coefficients [ "tx" ] )

# ops y
R177 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R177)
est2 [i, 177] <-  exp ( R177 $ coefficients [ "tx" ] )

# social y
R178 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R178)
est2 [i, 178] <-  exp ( R178 $ coefficients [ "tx" ] )



### Cut at rlb = .4 and quadratic weight
cut <- .4

# arb y
R179 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R179)
est2 [i, 179] <-  exp ( R179 $ coefficients [ "tx" ] )

# ops y
R180 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R180)
est2 [i, 180] <-  exp ( R180 $ coefficients [ "tx" ] )

# social y
R181 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R181)
est2 [i, 181] <-  exp ( R181 $ coefficients [ "tx" ] )



### Cut at rlb = .35 and quadratic weight
cut <- .35

# arb y
R182 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R182)
est2 [i, 182] <-  exp ( R182 $ coefficients [ "tx" ] )

# ops y
R183 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R183)
est2 [i, 183] <-  exp ( R183 $ coefficients [ "tx" ] )

# social y
R184 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R184)
est2 [i, 184] <-  exp ( R184 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .3 and quadratic weight
cut <- .3

# arb y
R185 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R185)
est2 [i, 185] <-  exp ( R185 $ coefficients [ "tx" ] )

# ops y
R186 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R186)
est2 [i, 186] <-  exp ( R186 $ coefficients [ "tx" ] )

# social y
R187 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R187)
est2 [i, 187] <-  exp ( R187 $ coefficients [ "tx" ] )



### Cut at rlb = .25 and quadratic weight
cut <- .25

# arb y
R188 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R188)
est2 [i, 188] <-  exp ( R188 $ coefficients [ "tx" ] )

# ops y
R189 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R189)
est2 [i, 189] <-  exp ( R189 $ coefficients [ "tx" ] )

# social y
R190 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R190)
est2 [i, 190] <-  exp ( R190 $ coefficients [ "tx" ] )



### Cut at rlb = .2 and quadratic weight
cut <- .2

# arb y
R191 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R191)
est2 [i, 191] <-  exp ( R191 $ coefficients [ "tx" ] )

# ops y
R192 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R192)
est2 [i, 192] <-  exp ( R192 $ coefficients [ "tx" ] )

# social y
R193 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R193)
est2 [i, 193] <-  exp ( R193 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .15 and quadratic weight
cut <- .15

# arb y
R194 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R194)
est2 [i, 194] <-  exp ( R194 $ coefficients [ "tx" ] )

# ops y
R195 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R195)
est2 [i, 195] <-  exp ( R195 $ coefficients [ "tx" ] )

# social y
R196 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R196)
est2 [i, 196] <-  exp ( R196 $ coefficients [ "tx" ] )



### Cut at rlb = .1 and quadratic weight
cut <- .1

# arb y
R197 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R197)
est2 [i, 197] <-  exp ( R197 $ coefficients [ "tx" ] )

# ops y
R198 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R198)
est2 [i, 198] <-  exp ( R198 $ coefficients [ "tx" ] )

# social y
R199 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R199)
est2 [i, 199] <-  exp ( R199 $ coefficients [ "tx" ] )



### Cut at rlb = .05 and quadratic weight
cut <- .05

# arb y
R200 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R200)
est2 [i, 200] <-  exp ( R200 $ coefficients [ "tx" ] )

# ops y
R201 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R201)
est2 [i, 201] <-  exp ( R201 $ coefficients [ "tx" ] )

# social y
R202 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^2,
sub = ( rlb >= cut ))
summary(R202)
est2 [i, 202] <-  exp ( R202 $ coefficients [ "tx" ] )
```

```
################# Threshold & Cubic Weight #################

### Cut at rlb = 1 and cubic weight
cut <- 1

# arb y
R203 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R203)
est2 [i, 203] <-  exp ( R203 $ coefficients [ "tx" ] )

# ops y
R204 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R204)
est2 [i, 204] <-  exp ( R204 $ coefficients [ "tx" ] )

# social y
R205 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R205)
est2 [i, 205] <-  exp ( R205 $ coefficients [ "tx" ] )


### Cut at rlb = .99 and cubic weight
cut <- .99

# arb y
R206 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R206)
est2 [i, 206] <-  exp ( R206 $ coefficients [ "tx" ] )

# ops y
R207 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R207)
est2 [i, 207] <-  exp ( R207 $ coefficients [ "tx" ] )

# social y
R208 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R208)
est2 [i, 208] <-  exp ( R208 $ coefficients [ "tx" ] )


### Cut at rlb = .95 and cubic weight
cut <- .95

# arb y
R209 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R209)
est2 [i, 209] <-  exp ( R209 $ coefficients [ "tx" ] )

# ops y
R210 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R210)
est2 [i, 210] <-  exp ( R210 $ coefficients [ "tx" ] )

# social y
R211 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R211)
est2 [i, 211] <-  exp ( R211 $ coefficients [ "tx" ] )

### Cut at rlb = .9 and cubic weight
cut <- .9

# arb y
R212 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R212)
est2 [i, 212] <-  exp ( R212 $ coefficients [ "tx" ] )

# ops y
R213 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R213)
est2 [i, 213] <-  exp ( R213 $ coefficients [ "tx" ] )

# social y
R214 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R214)
est2 [i, 214] <-  exp ( R214 $ coefficients [ "tx" ] )


### Cut at rlb = .85 and cubic weight
cut <- .85

# arb y
R215 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R215)
est2 [i, 215] <-  exp ( R215 $ coefficients [ "tx" ] )

# ops y
R216 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R216)
est2 [i, 216] <-  exp ( R216 $ coefficients [ "tx" ] )

# social y
R217 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R217)
est2 [i, 217] <-  exp ( R217 $ coefficients [ "tx" ] )


### Cut at rlb = .8 and cubic weight
cut <- .8

# arb y
R218 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R218)
est2 [i, 218] <-  exp ( R218 $ coefficients [ "tx" ] )

# ops y
R219 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R219)
est2 [i, 219] <-  exp ( R219 $ coefficients [ "tx" ] )

# social y
R220 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R220)
est2 [i, 220] <-  exp ( R220 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .75 and cubic weight
cut <- .75

# arb y
R221 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R221)
est2 [i, 221] <-  exp ( R221 $ coefficients [ "tx" ] )

# ops y
R222 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R222)
est2 [i, 222] <-  exp ( R222 $ coefficients [ "tx" ] )

# social y
R223 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R223)
est2 [i, 223] <-  exp ( R223 $ coefficients [ "tx" ] )


### Cut at rlb = .7 and cubic weight
cut <- .7

# arb y
R224 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R224)
est2 [i, 224] <-  exp ( R224 $ coefficients [ "tx" ] )

# ops y
R225 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R225)
est2 [i, 225] <-  exp ( R225 $ coefficients [ "tx" ] )

# social y
R226 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R226)
est2 [i, 226] <-  exp ( R226 $ coefficients [ "tx" ] )


### Cut at rlb = .65 and cubic weight
cut <- .65

# arb y
R227 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R227)
est2 [i, 227] <-  exp ( R227 $ coefficients [ "tx" ] )

# ops y
R228 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R228)
est2 [i, 228] <-  exp ( R228 $ coefficients [ "tx" ] )

# social y
R229 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R229)
est2 [i, 229] <-  exp ( R229 $ coefficients [ "tx" ] )

### Cut at rlb = .6 and cubic weight
cut <- .6

# arb y
R230 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R230)
est2 [i, 230] <-  exp ( R230 $ coefficients [ "tx" ] )

# ops y
R231 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R231)
est2 [i, 231] <-  exp ( R231 $ coefficients [ "tx" ] )

# social y
R232 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R232)
est2 [i, 232] <-  exp ( R232 $ coefficients [ "tx" ] )


### Cut at rlb = .55 and cubic weight
cut <- .55

# arb y
R233 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R233)
est2 [i, 233] <-  exp ( R233 $ coefficients [ "tx" ] )

# ops y
R234 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R234)
est2 [i, 234] <-  exp ( R234 $ coefficients [ "tx" ] )

# social y
R235 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R235)
est2 [i, 235] <-  exp ( R235 $ coefficients [ "tx" ] )


### Cut at rlb = .5 and cubic weight
cut <- .5

# arb y
R236 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R236)
est2 [i, 236] <-  exp ( R236 $ coefficients [ "tx" ] )

# ops y
R237 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R237)
est2 [i, 237] <-  exp ( R237 $ coefficients [ "tx" ] )

# social y
R238 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R238)
est2 [i, 238] <-  exp ( R238 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .45 and cubic weight
cut <- .45

# arb y
R239 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R239)
est2 [i, 239] <-  exp ( R239 $ coefficients [ "tx" ] )

# ops y
R240 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R240)
est2 [i, 240] <-  exp ( R240 $ coefficients [ "tx" ] )

# social y
R241 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R241)
est2 [i, 241] <-  exp ( R241 $ coefficients [ "tx" ] )


### Cut at rlb = .4 and cubic weight
cut <- .4

# arb y
R242 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R242)
est2 [i, 242] <-  exp ( R242 $ coefficients [ "tx" ] )

# ops y
R243 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R243)
est2 [i, 243] <-  exp ( R243 $ coefficients [ "tx" ] )

# social y
R244 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R244)
est2 [i, 244] <-  exp ( R244 $ coefficients [ "tx" ] )


### Cut at rlb = .35 and cubic weight
cut <- .35

# arb y
R245 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R245)
est2 [i, 245] <-  exp ( R245 $ coefficients [ "tx" ] )

# ops y
R246 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R246)
est2 [i, 246] <-  exp ( R246 $ coefficients [ "tx" ] )

# social y
R247 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R247)
est2 [i, 247] <-  exp ( R247 $ coefficients [ "tx" ] )

### Cut at rlb = .3 and cubic weight
cut <- .3

# arb y
R248 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R248)
est2 [i, 248] <-  exp ( R248 $ coefficients [ "tx" ] )

# ops y
R249 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R249)
est2 [i, 249] <-  exp ( R249 $ coefficients [ "tx" ] )

# social y
R250 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R250)
est2 [i, 250] <-  exp ( R250 $ coefficients [ "tx" ] )


### Cut at rlb = .25 and cubic weight
cut <- .25

# arb y
R251 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R251)
est2 [i, 251] <-  exp ( R251 $ coefficients [ "tx" ] )

# ops y
R252 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R252)
est2 [i, 252] <-  exp ( R252 $ coefficients [ "tx" ] )

# social y
R253 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R253)
est2 [i, 253] <-  exp ( R253 $ coefficients [ "tx" ] )


### Cut at rlb = .2 and cubic weight
cut <- .2

# arb y
R254 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R254)
est2 [i, 254] <-  exp ( R254 $ coefficients [ "tx" ] )

# ops y
R255 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R255)
est2 [i, 255] <-  exp ( R255 $ coefficients [ "tx" ] )

# social y
R256 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R256)
est2 [i, 256] <-  exp ( R256 $ coefficients [ "tx" ] )
```

```
### Cut at rlb = .15 and cubic weight
cut <- .15

# arb y
R257 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R257)
est2 [i, 257] <-  exp ( R257 $ coefficients [ "tx" ] )

# ops y
R258 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R258)
est2 [i, 258] <-  exp ( R258 $ coefficients [ "tx" ] )

# social y
R259 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R259)
est2 [i, 259] <-  exp ( R259 $ coefficients [ "tx" ] )



### Cut at rlb = .1 and cubic weight
cut <- .1

# arb y
R260 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R260)
est2 [i, 260] <-  exp ( R260 $ coefficients [ "tx" ] )

# ops y
R261 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R261)
est2 [i, 261] <-  exp ( R261 $ coefficients [ "tx" ] )

# social y
R262 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R262)
est2 [i, 262] <-  exp ( R262 $ coefficients [ "tx" ] )



### Cut at rlb = .05 and cubic weight
cut <- .05

# arb y
R263 <-  glm ( yy1 ~ tx + predictor1 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R263)
est2 [i, 263] <-  exp ( R263 $ coefficients [ "tx" ] )

# ops y
R264 <-  glm ( yy2 ~ tx + predictor2 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R264)
est2 [i, 264] <-  exp ( R264 $ coefficients [ "tx" ] )

# social y
R265 <-  glm ( yy3 ~ tx + predictor3 , fam = binomial , weight = rlb^3,
sub = ( rlb >= cut ))
summary(R265)
est2 [i, 265] <-  exp ( R265 $ coefficients [ "tx" ] )        }
```

Appendix C: R Code for *Bias* and *RRMSE* (%)

```
or <- -1.4204

sum.out <- matrix ("", 89, 4 )
sum.out [1,  ] <- c ( "", "Arbitrary", "Opposite", "Social Desirability" )
sum.out [, 1 ] <- c ( "", "Non-adjusted",
            "Cut at 1", "Cut at .99", "Cut at .95", "Cut at .9", "Cut at .85", "Cut at .8", "Cut at .75",
            "Cut at .7", "Cut at .65", "Cut at .6", "Cut at .55", "Cut at .5","Cut at .45", "Cut at .4",
            "Cut at .35", "Cut at .3", "Cut at .25", "Cut at .2", "Cut at .15", "Cut at .1", "Cut at .05",
            "Lin-weight", "Qua-weight", "Cub-weight",

            "Cut at 1 & LW", "Cut at .99 & LW", "Cut at .95 & LW", "Cut at .9 & LW", "Cut at .85 & LW", "Cut at .8 & LW", "Cut at .75 & LW",
            "Cut at .7 & LW", "Cut at .65 & LW", "Cut at .6 & LW", "Cut at .55 & LW", "Cut at .5 & LW", "Cut at .45 & LW", "Cut at .4 & LW",
            "Cut at .35 & LW", "Cut at .3 & LW", "Cut at .25 & LW", "Cut at .2 & LW", "Cut at .15 & LW", "Cut at .1 & LW", "Cut at .05 & LW",

            "Cut at 1 & QW", "Cut at .99 & QW", "Cut at .95 & QW", "Cut at .9 & QW", "Cut at .85 & QW", "Cut at .8 & QW", "Cut at .75 & QW",
            "Cut at .7 & QW", "Cut at .65 & QW", "Cut at .6 & QW", "Cut at .55 & QW", "Cut at .5 & QW", "Cut at .45 & QW", "Cut at .4 & QW",
            "Cut at .35 & QW", "Cut at .3 & QW", "Cut at .25 & QW", "Cut at .2 & QW", "Cut at .15 & QW", "Cut at .1 & QW", "Cut at .05 & QW",

            "Cut at 1 & CW", "Cut at .99 & CW", "Cut at .95 & CW", "Cut at .9 & CW", "Cut at .85 & CW", "Cut at .8 & CW", "Cut at .75 & CW",
            "Cut at .7 & CW", "Cut at .65 & CW", "Cut at .6 & CW", "Cut at .55 & CW", "Cut at .5 & CW", "Cut at .45 & CW", "Cut at .4 & CW",
            "Cut at .35 & CW", "Cut at .3 & CW", "Cut at .25 & CW", "Cut at .2 & CW", "Cut at .15 & CW", "Cut at .1 & CW", "Cut at .05 & CW" )

for ( i in (1:88 ) )            {
            for ( j in (1:3) )   {
            x <- log ( est2 [, 3 * (i-1) + j + 1 ])
            sum.out [ i + 1, j + 1 ] <- paste ( round (mean ( x )-or, 3 ), ", (", round (sqrt ( var ( x ) +  ( mean ( x ) - or ) ^ 2 )/abs(or) *100 , 3 ), ")" )
}
}

sum.out2 <- matrix ("", 92, 4 )
sum.out2 [ c(2:90), ] <- sum.out
sum.out2 [ 1, 1 ] <- "Compare Weighting Methods With Respect to Bias and RRMSE (%)"
sum.out2 [ 1, 2 ] <- paste (" n = ", n )
sum.out2 [ 1, 3 ] <- paste (" rep = ", rep )
sum.out2 [ 92, 1 ] <-"*If we had observed the true responses: "
x <- log (est2 [, 1])
sum.out2 [ 92, 2 ] <-paste ( round (mean ( x )-or, 3 ), ", (", round (sqrt ( var ( x ) +  ( mean ( x ) - or ) ^ 2 ) / abs(or)*100, 3 ), ")" )
sum.out2


###################### Export Results ######################


write.xlsx (sum.out2, file = "Simulation.xlsx", sheetName = paste("n=", n, sep="") , col.names = TRUE, row.names = TRUE, append = TRUE)
```

Appendix D: R Code for *RRMSE* (%) Figures

```
################################ Competing Bias Correction Strategies ##################################

library (xlsx)
library (ggplot2)

d <- read.xlsx("C:\\Users\\vswai\\Documents\\School\\MSPH Biostats\\Thesis\\Data\\Data.xlsx",sheetName = "Competing Strategies",
header=T)

summary(d$Bias)
summary(d$RRMSE)

d$Srlb2 <- factor(d$Srlb, levels = c("50", "80"))
d$Srlb2

levels (d$Srlb2) <- c("50%", "80%")
d$Srlb2

d$ES2 <- factor(d$ES, levels = c("Small", "Medium", "Large"))
d$ES2

# Data subset for RS = Opposite
drsopp <- d[d$RS=="Opposite",]
head(drsopp)
dim(drsopp)

# Data subset for RS = Arbitrary
drsarb <- d[d$RS=="Arbitrary",]
head(drsarb)
dim(drsarb)

# Data subset for RS = Social Desirability
drssd <- d[d$RS=="Social Desirability",]
head(drssd)
dim(drssd)

#Assign colors to specific bias-correction strategies
            #Mediumvioletred    = recommended strategy (N=200, 2000, & 20000)
            #Orange2            = alternate strategy 1 (N=200, 2000, & 20000)
            #Royalblue1         = alternate strategy 2 (N=2000 & 20000)
            #Mediumseagreen     = alternate strategy 3 (N=20000)
#Colors separated by N size (200, 2000, 20000)

color <- c( "S - Unadjusted*" = "Mediumvioletred", "S - LW" = "Orange2",
            "M - Cut at .40 & LW*" = "Mediumvioletred", "M - QW" = "Orange2",
            "M - Unadjusted*" = "Mediumvioletred",
            "L - Cut at .40 & QW*" = "Mediumvioletred", "L - CW" = "Orange2",
            "L - Unadjusted*" = "Mediumvioletred",

            "S - Cut at .40 & LW*" = "Mediumvioletred", "S - Cut at .25 & QW" = "Orange2",
            "S - QW" = "Royalblue1",
            "M - Cut at .55 & CW*" = "Mediumvioletred", "M - Cut at .65 & QW" = "Orange2",
            "M - Cut at .70 & LW" = "Royalblue1",
            "L - Cut at .75 & CW*" = "Mediumvioletred", "L - Cut at .80 & QW" = "Orange2",
            "L - Cut at .80 & LW" = "Royalblue1",

            "S - Cut at .70 & CW*" = "Mediumvioletred", "S - Cut at .85 & QW" = "Orange2",
            "S - Cut at .85 & LW" = "Royalblue1", "S - Cut at .85" = "Mediumseagreen",
            "M - Cut at .85 & CW*" = "Mediumvioletred", "M - Cut at .90 & QW" = "Orange2",
            "M - Cut at .90 & LW" = "Royalblue1", "M - Cut at .90" = "Mediumseagreen",
            "L - Cut at .95 & CW*" = "Mediumvioletred", "L - Cut at .95 & QW" = "Orange2",
            "L - Cut at .95 & LW" = "Royalblue1", "L - Cut at .95" = "Mediumseagreen" )
```

```
#Organize legend by (1) color (2) ES (3) Alpha-Numeric order
#Order of color: Mediumvioletred, Orange2, Royalblue1, Mediumseagreen

order <- c( "S - Unadjusted*", "S - Cut at .40 & LW*", "S - Cut at .70 & CW*",
            "M - Unadjusted*", "M - Cut at .40 & LW*", "M - Cut at .55 & CW*", "M - Cut at .85 & CW*",
            "L - Unadjusted*", "L - Cut at .40 & QW*", "L - Cut at .75 & CW*", "L - Cut at .95 & CW*",

            "S - LW", "S - Cut at .25 & QW", "S - Cut at .85 & QW",
            "M - QW", "M - Cut at .65 & QW", "M - Cut at .90 & QW",
            "L - CW", "L - Cut at .80 & QW", "L - Cut at .95 & QW",

            "S - QW", "S - Cut at .85 & LW",
            "M - Cut at .70 & LW", "M - Cut at .90 & LW",
            "L - Cut at .80 & LW", "L - Cut at .95 & LW",

            "S - Cut at .85",
            "M - Cut at .90",
            "L - Cut at .95" )



######## Comparison of RRMSE (%) for Competing Bias Correction Strategies ########

### RS = Opposite ### - FIGURE 3

p1 <- ggplot(drsopp, aes(Srlb2, RRMSE, group=Strategy, color=Strategy))  + geom_point (se=F, size=3) + geom_line (se=F)

p1 + scale_colour_manual(values=color, breaks=order) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Opposite Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))

### RS = Arbitrary ### - FIGURE 4

p2 <- ggplot(drsarb, aes(Srlb2, RRMSE, group=Strategy, color=Strategy))  + geom_point (se=F, size=3) + geom_line (se=F)

p2 + scale_colour_manual(values=color, breaks=order) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Arbitrary Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))

### RS = Social Desirability ### - FIGURE 5

p3 <- ggplot(drssd, aes(Srlb2, RRMSE, group=Strategy, color=Strategy))  + geom_point (se=F, size=3) + geom_line (se=F)

p3 + scale_colour_manual(values=color, breaks=order) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Socially Desirable Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))



################################  Undjusted and Adjusted Models  ##################################

d2 <- read.xlsx("C:\\Users\\vswai\\Documents\\School\\MSPH Biostats\\Thesis\\Data\\Data.xlsx",sheetName = "Unadjusted v Adjusted",
header=T)

summary(d2$Bias)
summary(d2$RRMSE)
```

```
d2$Srlb2 <- factor(d2$Srlb, levels = c("50", "80"))
d2$Srlb2

levels (d2$Srlb2) <- c("50%", "80%")
d2$Srlb2

d2$ES2 <- factor(d2$ES, levels = c("Small", "Medium", "Large"))
d2$ES2

d2$Adjustment <- factor(d2$Adj, levels = c("Unadjusted", "Recommended"))
d2$Adjustment

# Data subset for RS = Opposite
drsopp <- d2[d2$RS=="Opposite",]
head(drsopp)

# Data subset for RS = Arbitrary
drsarb <- d2[d2$RS=="Arbitrary",]
head(drsarb)

# Data subset for RS = Social Desirability
drssd <- d2[d2$RS=="Social Desirability",]
head(drssd)

#Assign colors to adjusted and unadjusted models
                #Mediumvioletred     = adjusted model (recommended strategy)
                #Orange2             = unadjusted model

color <- c( "Recommended" = "Mediumvioletred", "Unadjusted" = "Orange2")



######## Comparison of RRMSE (%) for Unadjusted and Adjusted Models ########

### RS = Opposite ### - FIGURE 6

p4 <- ggplot(drsopp, aes(Srlb2, RRMSE, group=Adjustment, color=Adjustment))  + geom_point (se=F, size=3) + geom_line (se=F)

p4 + scale_colour_manual(values=color) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Opposite Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))

### RS = Arbitrary ### - FIGURE 7

p5 <- ggplot(drsarb, aes(Srlb2, RRMSE, group=Adjustment, color=Adjustment))  + geom_point (se=F, size=3) + geom_line (se=F)

p5 + scale_colour_manual(values=color) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Arbitrary Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))

### RS = Social Desirability ### - FIGURE 8

p6 <- ggplot(drssd, aes(Srlb2, RRMSE, group=Adjustment, color=Adjustment))  + geom_point (se=F, size=3) + geom_line (se=F)

p6 + scale_colour_manual(values=color) +
        facet_grid(ES2~N) + labs(x="Sample Reliability", y="RRMSE (%)",
        title="Socially Desirable Responding") +
        theme(legend.position = "bottom") +
        theme(plot.title = element_text(hjust = 0.5)) +
        guides(color = guide_legend(title.position = "top", title.hjust = 0.5))
```