Sensitivity Analysis for Incomplete Data and Causal Inference

Heng Chen
hengc@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_statisticalscience_etds

Part of the Biostatistics Commons, Clinical Trials Commons, Statistical Methodology Commons, and the Statistical Theory Commons

Recommended Citation
https://scholar.smu.edu/hum_sci_statisticalscience_etds/14

This Thesis is brought to you for free and open access by the Statistical Science at SMU Scholar. It has been accepted for inclusion in Statistical Science Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
SENSITIVITY ANALYSIS FOR INCOMPLETE DATA AND CAUSAL INFERENCE

A Dissertation Presented to the Graduate Faculty of the
Dedman College
Southern Methodist University
in
Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy
with a
Major in Statistical Science
by
Heng Chen

B.S., Statistics, Tongji University

May 16, 2020
ACKNOWLEDGMENTS

First of all, I would like to express my special thanks of gratitude to Dr. Daniel Heitjan, who has guided, supported and inspired me in how to conduct scientific research during my PhD career; he continually and persuasively conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his supervision and constant help this dissertation would not have been possible.

Secondly, I would like to thank all my committee members, Dr. Sherry Wang, Dr. Jing Cao and Dr. Yin Xi, for their time and patience to help me greatly improve my dissertation. Furthermore, I would also like to acknowledge with much appreciation all the faculty members in Department of Statistical Science at SMU for their excellent teaching and help.

In addition, I would also like to express my deepest appreciation to all my friends during my PhD study, who have always been there to help me when I was struggling and comfort me when I was depressed.

Finally, my parents and family are always the most important people to me. I would like to thank them for their support and help. I will be grateful forever for them.
In this dissertation, we explore sensitivity analyses under three different types of incomplete data problems, including missing outcomes, missing outcomes and missing predictors, potential outcomes in *Rubin causal model (RCM)*. The first sensitivity analysis is conducted for the *missing completely at random (MCAR)* assumption in frequentist inference; the second one is conducted for the *missing at random (MAR)* assumption in likelihood inference; the third one is conducted for one novel assumption, the “sixth assumption” proposed for the robustness of instrumental variable estimand in causal inference.

In Chapter 2, we present a method to analyze sensitivity of frequentist inferences to potential nonignorability of the missingness mechanism. Rather than starting from the selection model, as is typical in such analyses, we assume that the missingness arises through unmeasured confounding. Our model permits the development of measures of sensitivity that are analogous to those for unmeasured confounding in observational studies. We define an index of sensitivity, denoted *MinNI*, to be the minimum degree of nonignorability needed to change the mean value of the estimate of interest by a designated amount. We apply our model to sensitivity analysis for a proportion, but the idea readily generalizes to more complex situations.

The ISNI (index of sensitivity to nonignorability) method quantifies local sensitivity of inferences to nonignorable missingness in an outcome variable. In Chapter 3, we extend the method to the situation where both outcomes and predictors can be missing. Ultimate
judgments about sensitivity rely on an evaluation of the minimum degree of nonignorability that gives rise to a defined, scientifically significant change in the estimate of a parameter of interest. We define the quantity $\text{MinNI (minimum nonignorability)}$ to be an approximation to the radius of the smallest ball centered at the MAR model in which nonignorability is negligible. We apply our method in a simulation study and two real-data examples involving the normal linear model and conditional logistic regression.

In Chapter 4, we explore the sensitivity of causal estimands in clinical trials with non-compliance. In a clinical trial with noncompliance, the selection of an estimand can be difficult. The intention-to-treat (ITT) analysis is a pragmatic approach, but the ITT estimand does not measure the causal effect of the actual treatment received and is sensitive to the level of compliance. An alternative estimand is the complier average causal effect (CACE), which refers to the average effect of treatment received in the latent subset of subjects who would comply with either treatment. Under the RCM, five assumptions are sufficient to identify CACE, permitting its consistent estimation from trial data. We observe that CACE can also vary with the fraction of compliance when the compliance class is regarded as a random quantity. We propose a “sixth assumption” that specifies that the individual-level compliance status and causal effect are independent in the super-population from which trial samples are drawn. This assumption guarantees robustness of CACE to the compliance fraction. We demonstrate the potential degree of sensitivity in a simulation study and an analysis of data from a trial of vitamin A supplementation in children. We observe that only CACE can be robust to varying levels of compliance, and only when the “sixth assumption” is satisfied.

In Chapter 5, we conclude our dissertation with further discussions for Chapter 2, Chapter 3 and Chapter 4.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. x

LIST OF TABLES ................................................................. xi

CHAPTER

1. INTRODUCTION .............................................................. 1
   1.1. Sensitivity analysis via unmeasured confounding .................. 1
   1.2. Local sensitivity analysis for missing outcomes and predictors .................. 3
   1.3. Sensitivity of estimands in clinical trials with noncompliance ........ 4

2. SENSITIVITY ANALYSIS VIA UNMEASURED CONFOUNDING ............ 6
   2.1. Model and methods ..................................................... 6
       2.1.1. Model and definitions ............................................ 6
       2.1.2. Sensitivity analysis in the confounding model .............. 8
   2.2. Response-surface sensitivity analysis ............................. 9
       2.2.1. Sensitivity parameters .......................................... 9
       2.2.2. Estimation of means with specified nonignorability parameters .... 10
   2.3. Identifying the minimum non-negligible nonignorability ............ 11
       2.3.1. MinNI in the difference scale ................................. 11
       2.3.2. MinNI in the ratio scale ....................................... 13
   2.4. Sensitivity analysis for the Edinburgh sexual behavior survey .......... 15
       2.4.1. The data .......................................................... 15
       2.4.2. A response-surface sensitivity analysis ..................... 15
       2.4.3. A MinNI sensitivity analysis compared with ISNI analysis ....... 16
       2.4.4. Dependence of MinNI on the fraction of missing data .......... 17
   2.5. Some extensions of the basic sensitivity analysis ................... 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1. A categorical confounder</td>
<td>17</td>
</tr>
<tr>
<td>2.5.2. Sensitivity analysis for the variance</td>
<td>19</td>
</tr>
<tr>
<td>2.5.3. Analysis with completely measured covariates</td>
<td>19</td>
</tr>
<tr>
<td>3. LOCAL SENSITIVITY ANALYSIS FOR MISSING OUTCOMES AND PREDICTORS</td>
<td>25</td>
</tr>
<tr>
<td>3.1. Methodology</td>
<td>25</td>
</tr>
<tr>
<td>3.1.1. ISNI</td>
<td>25</td>
</tr>
<tr>
<td>3.1.2. Interpretation of ISNI</td>
<td>28</td>
</tr>
<tr>
<td>3.2. Conditional Logistic Model</td>
<td>30</td>
</tr>
<tr>
<td>3.2.1. Conditional likelihood</td>
<td>30</td>
</tr>
<tr>
<td>3.2.2. Model specification</td>
<td>32</td>
</tr>
<tr>
<td>3.3. Simulated Missing Observations in the Smoking and Mortality Data</td>
<td>33</td>
</tr>
<tr>
<td>3.4. Real-Data Examples</td>
<td>35</td>
</tr>
<tr>
<td>3.4.1. The New York School Choice Experiment</td>
<td>35</td>
</tr>
<tr>
<td>3.4.2. The Los Angeles Endometrial Cancer Case Control Study</td>
<td>36</td>
</tr>
<tr>
<td>4. SENSITIVITY OF ESTIMANDS IN CLINICAL TRIALS WITH NONCOMPLIANCE</td>
<td>39</td>
</tr>
<tr>
<td>4.1. Clinical Trial Estimands</td>
<td>39</td>
</tr>
<tr>
<td>4.1.1. The RCM: Notation</td>
<td>39</td>
</tr>
<tr>
<td>4.1.2. Trial estimands viewed in light of the RCM</td>
<td>41</td>
</tr>
<tr>
<td>4.1.3. The sixth assumption</td>
<td>42</td>
</tr>
<tr>
<td>4.1.3.1. Definition</td>
<td>42</td>
</tr>
<tr>
<td>4.1.3.2. Relationship to other assumptions</td>
<td>44</td>
</tr>
<tr>
<td>4.2. Simulations</td>
<td>45</td>
</tr>
<tr>
<td>4.2.1. A simple model relating compliance to outcome</td>
<td>45</td>
</tr>
<tr>
<td>4.2.2. Model with always-takers</td>
<td>47</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 The equal-bias plot of ED$<em>{YU}$ and RD$</em>{UG}$ for the sexual behavior survey data. The numbers on the curves denote the bias in standard error units, with corresponding MinNI values (left to right) (0.10, 0.10), (0.14, 0.14), (0.20, 0.20), (0.24, 0.24), (0.28, 0.28), (0.31, 0.31), and (0.34, 0.34).</td>
<td>22</td>
</tr>
<tr>
<td>2.2 The equal-bias plot of ER$<em>{YU}$ and RR$</em>{UG}$ for the sexual behavior survey data. The numbers on the curves denote the bias in standard error units, with corresponding MinNI values (left to right) (1.13, 1.13), (1.19, 1.19), (1.30, 1.30), (1.39, 1.39), (1.48, 1.48), (1.56, 1.56), and (1.65, 1.65).</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Isobols of E[Y] - E[Y</td>
<td>G = 1] in terms of $\gamma_1$ and $\beta_1$, fixing $\pi_0 = 0.5$.</td>
</tr>
<tr>
<td>4.1 CACE and ITT$^{(Y)}$ as functions of $\rho$ and the proportion of compliers for fixed $\tau = 2$.</td>
<td>50</td>
</tr>
<tr>
<td>4.2 CACE and ITT$^{(Y)}$ as functions of $A_1$ and $A_2$ for fixed $\tau = 2$. (1) CACE with $\rho = 0.6$; (2) ITT$^{(Y)}$ with $\rho = 0.6$; (3) CACE with $\rho = 0$; (4) ITT$^{(Y)}$ with $\rho = 0$.</td>
<td>51</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>E[Y] − E[Y</td>
</tr>
<tr>
<td>2.2</td>
<td>E[Y] − E[Y</td>
</tr>
<tr>
<td>2.3</td>
<td>The MinNI giving one standard error bias with different fractions of missing data.</td>
</tr>
<tr>
<td>3.1</td>
<td>Sensitivity analysis for the slope in the smoking data, with artificially deleted data.</td>
</tr>
<tr>
<td>3.2</td>
<td>Missingness Patterns in the New York School Choice Experiment Data</td>
</tr>
<tr>
<td>3.3</td>
<td>Sensitivity Analysis for the New York School Choice Experiment</td>
</tr>
<tr>
<td>3.4</td>
<td>Sensitivity Analysis for the LA Endometrial Cancer Study</td>
</tr>
<tr>
<td>4.1</td>
<td>The Sommer-Zeger vitamin A supplement data</td>
</tr>
<tr>
<td>4.2</td>
<td>ITT(^{(Y)}) and CACE as a function of the proportion of compliers, (1 − \Phi(A_1)).</td>
</tr>
</tbody>
</table>
I dedicate this dissertation to my family and friends.
Incomplete data problems are encountered under various experiments, such as missing responses in surveys, missing outcomes or predictors in observational studies and noncompliance in randomized clinical trials. Each incomplete variable will induce one incompleteness mechanism; that is, the conditional distribution of the incompleteness indicators given the notional complete data. The ignorability conditions under which the stochastic nature of incompleteness mechanism could be ignored are of interest. However, these conditions are hard or impossible to verify from the observed data and then, sensitivity analysis is one simple and appealing approach to assess the robustness of inferences when the ignorability conditions are violated.

1.1. Sensitivity analysis via unmeasured confounding

Rubin has elucidated the role of the missingness mechanism in extracting frequentist inferences from incomplete data [51]. The idea is to compare the distribution of the variables that are observed, conditional on the observed missingness indicators, to the marginal distribution of these same variables ignoring the missingness mechanism. The condition missing completely at random (MCAR) is sufficient to guarantee that these distributions are identical, and therefore that the missingness mechanism is ignorable [36]. Briefly, MCAR requires that the conditional probability of the observed missingness indicators given the notional complete data is independent of the value of the complete data. Heitjan has extended this analysis to the coarse data model, where MCAR generalizes to
coarsened completely at random [22, 23].

The Rubin approach begins with a selection model; that is, it parameterizes the joint distribution of the variable of interest $Y$ and the missingness indicator $G$ as the product of the marginal distribution of $Y$ times the conditional distribution of $G$ given $Y$, denoting the latter the missingness mechanism. One can then identify ignorability conditions as restrictions on the parameters of the missingness mechanism. It is not possible to estimate the parameters of this joint model without strong assumptions [11]. An alternative, less ambitious, approach is to conduct sensitivity analyses to evaluate the robustness of estimates created under an ignorable model by re-estimating these parameters under a range of assumptions about the nonignorability parameters [9, 38, 58, 60].

In the selection model, one describes the probability of missingness as a function of the potentially missing observation; if the missingness and the outcomes are correlated (conventionally, if a nonignorability parameter is nonzero), ignorability does not hold. As a practical matter, it may be preferable to consider the correlation to arise from confounding — as indeed may be the case — in that both the outcomes and the missingness indicators are associated with a third variable. If we can identify and measure this variable, a form of conditional independence holds that guarantees ignorability. If we cannot, we posit a form for it and consider the consequences of nonignorability on the distributions of measurable outcomes. Such models have long served as a basis for sensitivity analysis in observational studies, where the concern is that the treatment indicators and the potential outcomes are correlated in a way that biases standard causal analyses [10, 13, 33, 41, 50, 59]. In Chapter 2, we apply this unconfounding specification to incomplete data.
1.2. Local sensitivity analysis for missing outcomes and predictors

A common model for data that are subject to missingness is the selection model. Rubin has elucidated conditions under which it is possible to ignore the stochastic nature of the missingness mechanism in Bayesian/likelihood inference [51]. Parameters distinctness (PD) asserts that the parameters governing the distribution of the notional complete data and the parameters governing the missingness mechanism lie in disjoint parameter spaces (for likelihood inference) or are a priori independent (for Bayesian inference). Missing at random (MAR) requires that the missingness mechanism does not depend on the values of the missing items. MAR and PD are sufficient to guarantee the ignorability of missingness mechanism in likelihood-based and Bayesian analyses. Analyses that ignore the missing-data mechanism therefore implicitly assume MAR and PD [19, 28–30, 34, 35, 45, 46, 52]. PD is often plausible, but MAR is generally not, and moreover it is impossible to verify by analyzing the available data.

A subset of the parameters of the missingness mechanism serve as nonignorability parameters, in that they govern the association of the complete data values and the missingness. Typically, we parameterize such models so that the missingness mechanism is MAR when ignorability parameters are set to 0. These parameters are impossible to estimate without strong assumptions, and even then, inferences may be numerically challenging and non-robust [11].

An alternative approach that is less ambitious but more practical is to evaluate the sensitivity of inferences to small departures from the MAR assumption; this is known as a local sensitivity analysis [9, 38, 58, 60, 64]. Such analyses to date have focused exclusively on situations where all missingness occurs in outcome variables, and none occurs in predictors. Moreover, in most such analyses there is a single nonignorability parameter, which considerably simplifies the interpretation of results. In Chapter 3, we extend an approach based on the index of local sensitivity to nonignorability (ISNI) [58]...
to the setting with both missing outcomes and predictors. In the process, we develop a flexible index, the \textit{minimum nonignorability (MinNI)}, for interpreting local sensitivity when there are multiple sources of nonignorability.

1.3. Sensitivity of estimands in clinical trials with noncompliance

Many clinical trials exhibit substantial fractions of noncompliance to assigned treatments, leading to the problem of whether and how to incorporate compliance in trial analysis. After many years of discussion, clinical trialists settled on the \textit{as-randomized (AR) or intention-to-treat (ITT)} analysis, which includes all subjects and groups them by the treatments to which they were randomized. The ITT estimate is unbiased for an estimand that represents the effect of random assignment, rather than treatment received, on the outcome. Advocates of this approach argue that it is more pragmatic than alternative analyses, as any attempts to assign treatments to a population will encounter some degree of noncompliance. A potential problem with ITT, however, is that it reflects the degree of noncompliance in the population under test; should compliance levels change — say, increasing over time to reflect growing recognition of a drug’s beneficial effects — the value of the estimand will also change.

Advocates of an approach to inference based on causal modeling have challenged this paradigm. A causal analysis takes a more structured approach to inference by first specifying the estimand of interest and then proceeding to identify conditions that render it estimable. For example, arguably the most important scientific estimand is the average effect of the treatment could all patients be compelled to comply with the randomization. Unfortunately, we cannot typically identify this estimand, known as the \textit{population average causal effect (PACE)}, from randomized experiments with noncompliance [3, 14, 25, 57]. An alternative estimand is the \textit{complier average causal effect (CACE)}, defined as the average effect of the treatment in the latent subset of subjects who would comply with
their randomization assignment, whatever it is. Under five plausible assumptions it is possible to identify and estimate CACE from outcome and compliance data [2].

A problem with both ITT and CACE is that they explicitly reflect patient compliance behavior, which can vary with many factors related to the nature of the treatment and the characteristics of the persons administering and receiving it [8, 14, 43, 44]. For example, a systematic review of patient compliance in clinical trials has demonstrated that compliance levels for a particular drug can range from 40% to 74% depending on dose frequency [43]. Investigators have shown that educational interventions can raise compliance for a hypertension treatment from 36% to 42% [44]. Thus, it is plausible that estimands that incorporate compliance can be unstable, in the sense of reflecting variation in compliance across formulations, populations, and time. In Chapter 4, we propose a novel assumption, “sixth assumption” to guarantee the robustness of CACE to randomness of noncompliance class and conduct sensitivity analyses in a simulation study and an illustrative real data example.
CHAPTER 2
SENSITIVITY ANALYSIS VIA UNMEASURED CONFOUNDING

In Section 2.1, we describe the model and establish the general ignorability conditions. Section 2.2 presents a response-surface method for assessing variation of parameters of interest as a function of nonignorability parameters in a parametric model. In Section 2.3, we adapt Cornfield’s paradigm, defining as an index of sensitivity the minimum magnitude of nonignorability that produces a designated level of bias. In Section 2.4 we apply the methods to incomplete data from a sexual behavior study. Section 2.5 covers extensions of the approach.

2.1. Model and methods

2.1.1. Model and definitions

The data consist of an outcome variable $Y = (Y_1, \ldots, Y_n)$ with corresponding vector of missingness indicators $G = (G_1, \ldots, G_n)$, where $G_i = 1$ for $Y_i$ observed, and $G_i = 0$ for $Y_i$ missing. Assume that an unmeasured variable $U = (U_1, \ldots, U_n)$ functions as confounder in that $Y$ and $G$ are conditionally independent given $U$; that is, the conditional distribution of $Y$ and $G$ given $U$ has the property that, for any $u$, $f_{Y,G|U}(y, g|u) = f_{Y|U}(y|u)f_{G|U}(g|u)$ for all $y$ and $g$. Thus, the joint density simplifies to

$$f_{Y,G,U}(y, g, u) = f_U(u)f_{Y|U}(y|u)f_{G|U}(g|u). \quad (2.1)$$
The confounding, if unmeasured or not accounted for, can induce correlation between $Y$ and $G$. Thus although we seek to create inferences for the marginal distribution of $Y$,

$$f^Y(y) = \int f^{Y|U}(y|u)f^U(u)du,$$

in fact we may be only able to observe the conditional distribution of $Y$ given $G = g$:

$$f^{Y|G}(y|g) = \int f^{G|U}(g|u)f^{Y|U}(y|u)f^U(u)du \int f^{G|U}(g|u)f^{Y|U}(y|u)f^U(u)dudy.$$

(2.3)

To this end, we establish restrictions on the conditional distribution terms in Equation (2.3) that are sufficient to guarantee ignorability, which in this context means that Equations (2.2) and (2.3) are the same. Throughout, we ignore any theoretical considerations about sets of measure 0.

**Theorem 2.1** Assume that either $G \perp \perp U$, or $Y \perp \perp U$. Then for any $g$ such that $f^G(g) > 0$, the distribution ignoring the missing mechanism in Equation (2.2) equals the correct distribution in Equation (2.3).

Proof. Suppose $G \perp \perp U$. Then $\forall g$ with $0 < f^G(g) < 1$,

$$f^{Y|G}(y|g) = \frac{\int f^{Y|U}(y|u)f^G(g)f^U(u)du}{\int f^{Y|U}(y|u)f^G(g)f^U(u)dudy} = \frac{\int f^{Y|U}(y|u)f^U(u)du}{\int f^{Y|U}(y|u)f^U(u)dudy} = f^Y(y).$$

Similarly, if $Y \perp \perp U$, then $\forall g$ with $0 < f^G(g) < 1$,

$$f^{Y|G}(y|g) = \frac{f^Y(y)\int f^{G|U}(g|u)f^U(u)du}{\int f^{G|U}(g|u)f^U(u)du} = f^Y(y).$$

If $Y$ or $U$ is discrete, one can restate the theorem with summation substituted for integration. The ignorability condition in the theorem is stronger than MCAR because it applies to all possible missing patterns that have positive density, not just the observed
missing pattern [23, 36]. In practice, the relevant conditional distribution will be the one for
the outcome \( y \) conditional on the observed vector of missingness indicators \( \tilde{g} \) where the
sample space of \( y \) will be restricted to those which could agree with \( \tilde{g} \), called \( y \) consistent
with \( \tilde{g} \). Thus, we develop an alternative, weaker version of Theorem 2.1.

Consider the following conditions, assuming an observed value \( \tilde{g} \) of \( g \):

1. The missingness is \textit{observed ignorable} in that for any possible \( u \), \( f^{G|Y,U}(\tilde{g}|y,u) \) takes
   the same value for all \( y \) consistent with \( \tilde{g} \).
2. \( f^{G|U}(\tilde{g}|u) \) takes the same value for all \( u \).
3. For any \( y \) consistent with \( \tilde{g} \), \( f^{Y|U}(y|u) \) takes the same value for all \( u \).

This leads to the following theorem:

\textbf{Theorem 2.2} \textit{Under Assumption 1 and either of Assumptions 2 or 3, \( f^Y(y) = f^{Y|G}(y|\tilde{g}) \)
for all \( y \) consistent with \( \tilde{g} \).}

In practice, there may also be completely measured predictors. In such a case the
theorems go through with appropriate conditioning, as shown in Section 2.5.3. These
theorems offer the simplest general ignorability conditions for the confounding model. Ig-
rorability is generally not testable because \( U \) is typically hypothetical and \( Y_i \) is available
only when \( G_i = 1 \). Our idea therefore is to define nonignorability parameters in the con-
text of Equation (2.1), then manipulate those parameters to determine how far they must
depart from the ignorable model to create a substantial difference between \( f^Y(y) \) and
\( f^{Y|G}(y|\tilde{g}) \).

2.1.2. Sensitivity analysis in the confounding model

To simplify our exposition, we consider the situation where the data represent \( n \) in-
dependently and identically distributed cases, with \( U \) a scalar unmeasured confounder.
Assuming that $Y$ has finite mean and variance, we base our sensitivity analysis initially on a comparison of the marginal mean of $Y$ to its mean conditional on its being observed.

Theorems 2.1 and 2.2 suggest that the sensitivity parameters can represent associations between $U$ and $Y$ and between $U$ and $G$. We will describe two approaches: In the first, we depict bias conventionally by varying the nonignorability parameters over a plausible range based on a mildly parameterized model. This extends the sensitivity analysis of Rosenbaum and Rubin [50] from confounding in observational studies to nonignorably missing data. In the second, we consider a minimally parameterized nonparametric model and define the minimum nonignorability index (MinNI) to be the degree of nonignorability necessary to cause a non-negligible bias. This is similar to the Cornfield approach [10] to sensitivity analysis in observational research. As with all sensitivity analyses, ours depends in principle on the judgments of a hypothetical expert, whose role it is to identify the minimum non-negligible values of both the bias in $Y$ and the nonignorability parameters.

2.2. Response-surface sensitivity analysis

2.2.1. Sensitivity parameters

Assume first that the confounder $U$ is binary. Then we partially specify the joint distribution of $(Y, G, U)$ as

$$\Pr[U = 0] = \pi_0,$$

$$\Pr[G = 1|U = u] = h(\gamma_0 + \gamma_1 u),$$

$$\mathbb{E}[Y|U = u] = q(\beta_0 + \beta_1 u),$$

where $u \in \{0, 1\}$ and $h(\gamma_0 + \gamma_1 u)$ and $q(\beta_0 + \beta_1 u)$ are link functions. The parameters $\pi_0$, $\gamma_1$, and $\beta_1$ describe the degree of sensitivity; they are unidentifiable because we do not
observe $U$.

A standard approach to sensitivity analysis is to observe the change in a parameter of interest, in this case the marginal mean $E[Y]$, as we vary the sensitivity parameters over plausible values. Under this model, the marginal and conditional means of $Y$ in terms of these sensitivity parameters are, respectively,

\begin{align}
E[Y] &= q(\beta_0 + \beta_1)(1 - \pi_0) + q(\beta_0)\pi_0 \\
E[Y|G = 1] &= \frac{q(\beta_0 + \beta_1)h(\gamma_0 + \gamma_1)(1 - \pi_0) + q(\beta_0)h(\gamma_0)\pi_0}{h(\gamma_0 + \gamma_1)(1 - \pi_0) + h(\gamma_0)\pi_0}
\end{align}

2.2.2. Estimation of means with specified nonignorability parameters

Under this model, $Pr[G = 1]$ and $E[Y|G = 1]$ are directly estimable from the data as $\hat{p}$ and $\hat{\mu}_c$, respectively. With $\pi_0$, $\gamma_1$, and $\beta_1$ fixed, and observing a random sample of $Y$ values, some of which may be missing, one can readily estimate $\gamma_0$ and $\beta_0$ [50]. The first estimable term is

\[ Pr[G = 1] = h(\gamma_0 + \gamma_1)(1 - \pi_0) + h(\gamma_0)\pi_0. \]

Thus we have two Equations ((2.5) and (2.6)) and two unknowns ($\gamma_0$ and $\beta_0$). We calculate the marginal mean in (2.4) as follows:

1. Solve Equation (2.6) for $\hat{\gamma}_0$, with $\hat{p}$ and $\pi_0, \gamma_1$ fixed;

2. Solve Equation (2.5) for $\beta_0$, with $\pi_0, \gamma_1, \beta_1$ fixed, $\hat{\gamma}_0$ from step 1, and $\hat{\mu}_c$ estimated directly from the data;

3. Substitute $\hat{\beta}_0$ and $\pi_0, \beta_1$ into (2.4) to estimate the marginal mean.

Appendix A.1 presents details for the special case where both link functions are logistic, as would be applicable with a binary outcome.
Theorems 2.1 and 2.2 assert that if $\gamma_1 = 0$ or $\beta_1 = 0$, there is no difference between $\mathbb{E}[Y]$ and $\mathbb{E}[Y|G = 1]$. Thus if small values of these parameters lead to substantial variation in $\mathbb{E}[Y]$, we deem the results sensitive. If the response surface for $\mathbb{E}[Y]$ as a function of the sensitivity parameters is flat, then only large values of the sensitivity parameters imply non-negligible changes in $\mathbb{E}[Y|G = 1]$, and inferences are insensitive.

If the notional unmeasured covariate $U$ is other than binary, the specification of the distribution for $U$ is more complex and may involve more parameters. The distributions of $G$ given $U$ and $Y$ given $U$ are indexed by link functions, whose specification induces an additional source of sensitivity. Thus, semiparametric or nonparametric models might be more satisfactory for this application.

2.3. Identifying the minimum non-negligible nonignorability

The response-surface analysis directly investigates the bias by mapping the effects of nonignorability on the distribution of $Y$. A complementary approach is to identify minimum values for the sensitivity parameters that yield a designated level of change — in this case, a pre-specified maximum negligible difference between $\mathbb{E}[Y|G = 1]$ and $\mathbb{E}[Y]$. We denote these parameter values $\text{MinNI}$, for Minimum NonIgnorability. We seek moreover to conduct the analysis with a minimally parameterized model.

2.3.1. MinNI in the difference scale

Assume again a binary confounder $U$. We first note that

$$\mathbb{E}[Y] - \mathbb{E}[Y|G = 1] = (\mathbb{E}[Y|G = 0] - \mathbb{E}[Y|G = 1]) \Pr[G = 0]. \quad (2.7)$$

Clearly, unless $0 < \Pr[G = 0] < 1$ there is no need for a sensitivity analysis. Expanding the bias in Equation (2.7) in terms of the unmeasured confounder $U$, we observe that the
difference between $E[Y|G = 0]$ and $E[Y|G = 1]$ can be decomposed into the product of
the difference between $E[Y|U = 1]$ and $E[Y|U = 0]$ and the difference between $Pr[U = 1|G = 1]$ and $Pr[U = 1|G = 0]$. Details appear in Appendix A.2.

Define the sensitivity parameters as the two differences

$$ED_{YU} = E[Y|U = 1] - E[Y|U = 0], \quad RD_{UG} = Pr[U = 1|G = 1] - Pr[U = 1|G = 0],$$

and observe that

$$|E[Y] - E[Y|G = 1]| = |ED_{YU}RD_{UG}Pr[G = 0]|. \quad (2.8)$$

We can construct an insensitive region by specifying a maximum negligible difference
for the bias as

$$|E[Y] - E[Y|G = 1]| \leq k\sigma_{Y|G = 1}, \quad (2.9)$$

where $\sigma_{Y|G = 1}$ is the standard deviation of $Y$ given it is observed, and $k$ is a positive con-
stant defined for the context, possibly related to sample size. From (2.8) and (2.9), we
obtain the indifference region for the nonignorable parameters to be

$$|ED_{YU}RD_{UG}| \leq \frac{k\sigma_{Y|G = 1}}{Pr[G = 0]}. \quad (2.10)$$

The Inequality (2.10) describes the relations among the maximum tolerable change and
the sensitivity parameters. To define a single index of sensitivity, we identify the combina-
tion of sensitivity parameters that satisfies this constraint and is closest to the origin. We
call this the MinNI for the mean. For a continuous outcome, the optimization process is

Minimize: \((ED_{YU}^2 + RD_{UG}^2)\)

Subject to:

\(|ED_{YU}| \cdot RD_{UG} \leq \frac{k\sigma_{Y|G=1}}{Pr[G=0]}\);

\(|ED_{YU}| \in (0, \infty)\);

\(|RD_{UG}| \in (0, 1)\).

The closed-form feasible solution (i.e. MinNI) for \((|ED_{YU}|, |RD_{UG}|)\) is

\[
\left( \max \left\{ \frac{k\sigma_{Y|G=1}}{Pr[G=0]}, \sqrt{\frac{k\sigma_{Y|G=1}}{Pr[G=0]}} \right\}, \min \left\{ 1, \sqrt{\frac{k\sigma_{Y|G=1}}{Pr[G=0]}} \right\} \right). 
\]

For a binary outcome, the range of \(|ED_{YU}|\) in the optimization procedure is \((0, 1)\), and

\[\text{MinNI} = \left( \sqrt{\frac{k\sigma_{Y|G=1}}{Pr[G=0]}}, \sqrt{\frac{k\sigma_{Y|G=1}}{Pr[G=0]}} \right),\]

where \(k\sigma_{Y|G=1} \leq Pr[G=0]\). If MinNI is large, the sampling inference ignoring the missing data is plausibly robust. If it is small, ignoring the missing mechanism could cause a considerable bias. Figure 2.1 illustrates the sensitivity analysis of the example discussed in Section 2.4 below.

2.3.2. MinNI in the ratio scale

For categorical variables, it might be preferable to describe bias on the ratio scale. Analogously with Equation (2.7), we observe that

\[
\frac{E[Y]}{E[Y|G=1]} = Pr[G=1] + Pr[G=0] \frac{E[Y|G=0]}{E[Y|G=1]},
\]

(2.11)
where $E[Y|G=1] \neq 0$. Defining the nonignorability parameters as the ratios

$$ER_{YU} = \frac{E[Y|U=1]}{E[Y|U=0]}; \quad RR_{UG} = \frac{P(U=1|G=1)}{P(U=1|G=0)},$$

we obtain

$$E[Y|G=0] = \frac{ER_{YU} - 1 + \frac{1}{Pr[U=1|G=0]}}{(ER_{YU} - 1)RR_{UG} + \frac{1}{Pr[U=1|G=0]}}.$$

(2.12)

Because we cannot identify $Pr[U=1|G=0]$, the best we can do is to obtain inequalities on the ratio in Equation (2.12), whose right-hand side is a monotone function of $\frac{1}{Pr[U=1|G=0]} \in (RR_{UG}, \infty)$. We express the bounding inequality for the original ratio as

$$\left| \frac{E[Y]}{E[Y|G=1]} - 1 \right| \leq \left| \frac{(ER_{YU} - 1)(RR_{UG} - 1)}{ER_{YU}RR_{UG}} \right| Pr[G=0],$$

(2.13)

where $ER_{YU} \in (-\infty, \infty)$ and $RR_{UG} \in (0, \infty)$. When $Y$ is binary, we can specify an indifference region on the ratio scale by dividing both sides in (2.9) by $E[Y|G=1]$ to obtain

$$\left| \frac{E[Y]}{E[Y|G=1]} - 1 \right| \leq |kCV_{Y|G=1}|.$$

(2.14)

Here $CV_{Y|G=1}$ is the coefficient of variation of $Y$ given that it is observed. The parameters $Pr[G=0], \sigma_{Y|G=1}$ and $CV_{Y|G=1}$ are all estimable from the data. To be conservative, we make the upper bound of the ratio in Inequality (2.13) less than the specified detectable difference from (2.14). The indifference region for the nonignorable ratio parameters is then

$$\left| \frac{(ER_{YU} - 1)(RR_{UG} - 1)}{ER_{YU}RR_{UG}} \right| \leq \frac{|kCV_{Y|G=1}|}{Pr[G=0]}.$$

(2.15)

To obtain a sensitivity index, we identify the closest point to $(1,1)$. Assuming, without loss
of generality, that both $ER_{YU}$ and $RR_{UG}$ exceed 1, the optimization process is

Minimize: \[(ER_{YU} - 1)^2 + (RR_{UG} - 1)^2\]
Subject to: \[\frac{(ER_{YU} - 1)(RR_{UG} - 1)}{ER_{YU}RR_{UG}} \leq \frac{|kCV_{Y|G=1}|}{Pr[G=0]};\]
\[ER_{YU} \in (1, \infty);\]
\[RR_{UG} \in (1, \infty).\]

The closed-form solution for $(ER_{YU}, RR_{UG})$ is then

\[
\text{MinNI} = \left(\frac{1}{1 - \sqrt{\frac{|kCV_{Y|G=1}|}{Pr[G=0]}}, \frac{1}{1 - \sqrt{\frac{|kCV_{Y|G=1}|}{Pr[G=0]}}}}\right),
\]

where $|kCV_{Y|G=1}| < Pr[G = 0] \leq 1$. The interpretation is the same as for the difference scale; see Figure 2.2.

2.4. Sensitivity analysis for the Edinburgh sexual behavior survey

2.4.1. The data

Investigators surveyed 6,136 randomly selected students at the University of Edinburgh in 1993. The parameter of main interest was the fraction responding “yes” to the question “Have you ever had sexual intercourse?”, which 2,308 students (37.6%) declined to answer [47, 58, 62]. The observed proportion of positive responses, estimating $E[Y|G = 1]$, is 0.7320 with standard error 0.0072. There is concern that nonresponders could have different patterns of sexual behavior compared to responders, potentially inducing a bias when estimating the parameter of interest. We describe below a sensitivity analysis for this proportion.
2.4.2. A response-surface sensitivity analysis

Table 2.1 displays the bias as a function of the sensitivity parameters, \( \pi_0, \beta_1, \) and \( \gamma_1 \). A plot of equal-bias contours in \( \gamma_1 \) and \( \beta_1 \), with \( \pi_0 \) fixed at 0.5, appears in Figure 2.3. We fixed \( \pi_0 = 0.5 \) because this value appears to give the largest bias.

In Table 2.1, the absolute magnitude of the bias is modest as a fraction of the estimated parameter, reaching values no larger than about 3% on a relative scale. For purposes of statistical inference, however, the sensitivity is substantial, as the largest bias is roughly 3 times the nominal standard error. The equal-bias plot in Figure 2.3 indicates that moderate values of \( \gamma_1 \) and \( \beta_1 \) can lead to 2-SE changes to the mean. The analysis thus suggests that estimation of the proportion of students who had had sexual intercourse is sensitive to nonignorability.

2.4.3. A MinNI sensitivity analysis compared with ISNI analysis

Here we set the maximum negligible bias to be 1 standard error of the observed proportion (here \( k\sigma_{Y|G=1} = 0.0072 \)) and compute minimum values of the sensitivity parameters that produce this level of displacement. The MinNI for the difference scale, \((ED_{YU}, RD_{UG})=(0.14,0.14)\) from Figure 2.1 and for the ratio scale \((ER_{YU}, RR_{UG})=(1.19,1.19)\) from Figure 2.2. The index is in both cases small, suggesting that the sampling inference for the true proportion of having sexual intercourse is sensitive. That is, even a modest disturbance from the ignorable model can induce a substantial bias into our estimate of the population proportion, rendering tests and confidence intervals for this parameter unreliable.

We compare this analysis with an application of the likelihood-based ISNI (index of local sensitivity to nonignorability) sensitivity analysis \[58, 62\]. With ISNI, the key sensitivity statistic, denoted \( c \), measures the approximate minimum standardized magnitude
of nonignorability needed to induce a 1-SE change in the maximum likelihood estimate of the parameter of interest. A value $c < 1$ is generally taken as evidence of sensitivity. For the proportion replying yes in the Edinburgh data, we compute $c = 0.097$, suggesting strong sensitivity and agreeing with our frequentist analysis.

2.4.4. Dependence of MinNI on the fraction of missing data

Measures of sensitivity to nonignorability depend critically on the fraction of missing data; indeed the ISNI measure for a univariate normal mean with missing observations is proportional to the fraction missing [58]. To illustrate this relationship, we artificially varied the fraction of missing observations while holding the observed fraction of positive responses constant. We repeated the analysis with artificial missingness fractions set to 0.1 and 0.2, both smaller than the observed value of 0.376. Table 2.2 shows the dependence of the bias for $E[Y]$ as a function of the response-surface sensitivity parameters. Recalling that the standard error of the observed fraction of responses is 0.0072, it is clear that for smaller fractions of missing data, sensitivity is modest except for the most extreme levels of confounding.

Table 2.3 shows MinNI values for the difference and ratio scale sensitivity analyses under the alternative fractions of missing observations. The interpretation of these values is that one would require weaker levels of confounding to induce a non-negligible bias in the observed fraction of positive responses.

2.5. Some extensions of the basic sensitivity analysis

2.5.1. A categorical confounder

It is straightforward to extend our analysis to the case of an unmeasured confounder
with \( m > 2 \) levels. For the difference scale, denote the confounding relations as follows:

\[
\text{MD}_{YU} = \max_i \mathbb{E}[Y|U = u_i] - \min_i \mathbb{E}[Y|U = u_i],
\]

\[
\text{MD}_{UG} = \max_i [\Pr[U = u_i|G = 1] - \Pr[U = u_i|G = 0]].
\]

In Appendix A.3.1 we derive the bounding inequality to be

\[
|\mathbb{E}[Y] - \mathbb{E}[Y|G = 1]| \leq |(m - 1)\text{MD}_{YU}\text{MD}_{UG}\Pr[G = 0]|. \tag{2.16}
\]

To be conservative, we make the upper bound of the difference less than \( k \) standard deviations of the observed standard deviation \( \sigma_{Y|G=1} \),

\[
|\text{MD}_{YU}\text{MD}_{UG}| \leq \frac{k\sigma_{Y|G=1}}{(m - 1)\Pr[G = 0]]. \tag{2.17}
\]

The dependence of the sensitivity on the number of categories is the same as found in Ding and VanderWeele (2014).

For the relative ratio scale, we denote the confounding parameters to be

\[
\text{ER}_{YU(i)} = \frac{\mathbb{E}[Y|U = u_i]}{\min_i \mathbb{E}[Y|U = u_i]}, \quad \text{RR}_{UG(i)} = \frac{\Pr[U = u_i|G = 1]}{\Pr[U = u_i|G = 0]},
\]

\[
\text{MR}_{YU} = \max_i \text{ER}_{YU(i)}, \quad \text{MR}_{UG} = \max_i \text{RR}_{UG(i)}.
\]

Without loss of generality, we can take all of these parameters to be greater than 1. In Appendix A.3.1 we show the bounding inequality to be

\[
\left| \frac{\mathbb{E}[Y]}{\mathbb{E}[Y|G = 1]} - 1 \right| \leq \left| \frac{(\text{MR}_{YU} - 1)(\text{MR}_{UG} - 1)}{\text{MR}_{YU}\text{MR}_{UG}} \right|. \tag{2.18}
\]
This leads to the conservative indifference region

\[
\frac{(MR_{UY} - 1)(MR_{UG} - 1)}{MR_{UY} MR_{UG}} \leq \frac{kCV_{Y|G=1}}{\Pr[G = 0]}.
\]  

(2.19)

The corresponding MinNI derivations appear in Appendix A.3.2.

2.5.2. Sensitivity analysis for the variance

So far we have only considered bias in the mean of \(Y\), but bias can also affect the variance. In an obvious notation, we define \(\sigma^2_B\) to be the variance of a random variable \(B\), potentially with conditioning. Setting

\[
VD_{UY} = \sigma^2_Y|U=0 - \sigma^2_Y|U=1, \quad VD_{UG} = \sigma^2_U|G=0 - \sigma^2_U|G=1,
\]

we obtain

\[
\sigma^2_Y - \sigma^2_Y|G=1 = \{VD_{UY}RD_{UG} + ED_{UY}^2VD_{UG} + ED_{UY}^2RD_{UG}^2Pr[G=1]\} Pr[G = 0]. \tag{2.20}
\]

Theorem 2.1 asserts that if \(G \perp \perp U\) or \(Y \perp \perp U\), then the difference in Equation (2.20) is 0. For the comparison of means, if either \(ED_{UY}\) or \(RD_{UG}\) is 0, there is no bias, but for the comparison of variance, this condition is not sufficient because \(VD_{UY}\) or \(VD_{UG}\) might not be 0. Commonly, the main moment of interest is the mean and it is shown that the first-order Taylor expansion of \(\sigma^2_Y|G=1\) is equal to \(\sigma^2_Y\) [37, 58]. We can readily derive analogous results for estimating the conditional distribution of \(Y\) given \(X\).

2.5.3. Analysis with completely measured covariates

Many studies will include many baseline variables that, if unobserved, would confound the association of outcome and missingness; we denote such variables \(X\). We can readily generalize Theorems 2.1 and 2.2 to cover estimation of the distribution of \(Y\) given \(X\).
**Theorem 2.3** Assume that $Y \perp \perp G \mid (X, U)$ and that either $G \perp \perp U \mid X$ or $Y \perp \perp U \mid X$. Then for any $g$ and $x$ such that $f^{G,X}(g, x) > 0$, the distribution ignoring the missing mechanism $f^{Y \mid X}(y \mid x)$, equals the correct distribution $f^{Y \mid G,X}(y \mid g, x)$.

To generalize Theorem 2.2, we define the following assumptions, assuming that $\tilde{g}$ is the observed value of $G$:

1. The missingness is observed ignorable in that for any possible $u$ and $x$,

   $$f^{G \mid Y,X,U}(\tilde{g} \mid y, x, u)$$

   takes the same value for all $y$ consistent with $\tilde{g}$.

2. For any possible $x$, $f^{G \mid X,U}(\tilde{g} \mid x, u)$ takes the same value for all $u$.

3. For any possible $x$ and any $y$ consistent with $\tilde{g}$, $f^{Y \mid X,U}(y \mid x, u)$ takes the same value for all $u$.

**Theorem 2.4** Under Assumption 1 and either of Assumptions 2 or 3,

$$f^{Y \mid X}(y \mid x) = f^{Y \mid X,G}(y \mid x, \tilde{g})$$

for all $y$ consistent with $\tilde{g}$.

Our analysis also readily generalizes to this situation; that is, by further conditioning on $X$ one can elucidate sensitivity as we have done above. Assume that the measured covariates $X$ are discrete. For the difference scale, denote the two confounding relations as

$$\text{ED}_{YU(X)} = E[Y \mid X, U = 1] - E[Y \mid X, U = 0],$$

$$\text{RD}_{UG(X)} = \Pr[U = 1 \mid X, G = 1] - \Pr[U = 1 \mid X, G = 0].$$
Therefore,

\[ |E[Y|X] - E[Y|X, G = 1]| = |ED_{YU(X)}RD_{U|G(X)}Pr[G = 0|X]|. \]

The above formula is similar to Equation (2.8). However, for the ratio scale, one naive analysis will be shown below.

\[ \frac{E[Y|X]}{E[Y|X, G = 1]} = Pr[G = 1|X] + \frac{E[Y|X, G = 0]}{E[Y|X, G = 1]} Pr[G = 0|X], \]

and we denote the relative ratios as

\[ ER_{YU(X)} = \frac{E[Y|X, U = 1]}{E[Y|X, U = 0]}, \]

\[ RR_{U|G(X)} = \frac{Pr[U = 1|X, G = 1]}{Pr[U = 1|X, G = 0]}. \]

Hence,

\[ \frac{E[Y|X, G = 0]}{E[Y|X, G = 1]} = \frac{ER_{YU(X)} - 1 + \frac{1}{Pr[U = 1|X, G = 0]}}{(ER_{YU(X)} - 1)RR_{U|G(X)} + \frac{1}{Pr[U = 1|X, G = 0]}}, \]

which recalls Equation (2.12). All the other derivations follow directly. The total discrepancy between the marginal mean and the conditional mean after adjusting for the \( X \) could be the summation of the discrepancy weighted by \( X \).
Figure 2.1: The equal-bias plot of ED$_{YU}$ and RD$_{UG}$ for the sexual behavior survey data. The numbers on the curves denote the bias in standard error units, with corresponding MinNI values (left to right) (0.10, 0.10), (0.14, 0.14), (0.20, 0.20), (0.24, 0.24), (0.28, 0.28), (0.31, 0.31), and (0.34, 0.34).

Table 2.1: $E[Y] - E[Y|G = 1]$ as a function of the sensitivity parameters.

<table>
<thead>
<tr>
<th>exp($\beta_1$)</th>
<th>exp($\gamma_1$)</th>
<th>$\pi_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>-0.0037</td>
<td>-0.0088</td>
</tr>
<tr>
<td>3</td>
<td>-0.0059</td>
<td>-0.0139</td>
</tr>
<tr>
<td>2</td>
<td>-0.0061</td>
<td>-0.0138</td>
</tr>
<tr>
<td>3</td>
<td>-0.0097</td>
<td>-0.0218</td>
</tr>
</tbody>
</table>
Figure 2.2: The equal-bias plot of ER_{YU} and RR_{UG} for the sexual behavior survey data. The numbers on the curves denote the bias in standard error units, with corresponding MinNI values (left to right) (1.13, 1.13), (1.19, 1.19), (1.30, 1.30), (1.39, 1.39), (1.48, 1.48), (1.56, 1.56), and (1.65, 1.65).

Table 2.2: E[Y] − E[Y|G = 1] as a function of the sensitivity parameters with \( \pi_0 = 0.5 \).

<table>
<thead>
<tr>
<th>( \exp(\beta_1) )</th>
<th>( \exp(\gamma_1) )</th>
<th>0.1</th>
<th>0.2</th>
<th>0.376</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.0023</td>
<td>-0.0046</td>
<td>-0.0088</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.0035</td>
<td>-0.0071</td>
<td>-0.0139</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.0035</td>
<td>-0.0072</td>
<td>-0.0138</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.0054</td>
<td>-0.0111</td>
<td>-0.0218</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.3: Isobols of $E[Y] - E[Y|G = 1]$ in terms of $\gamma_1$ and $\beta_1$, fixing $\pi_0 = 0.5$.

Table 2.3: The MinNI giving one standard error bias with different fractions of missing data.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Fraction missing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>$(</td>
<td>ED_{YU}</td>
</tr>
<tr>
<td>$(ER_{YU},RR_{UG})$</td>
<td>(1.46,1.46)</td>
</tr>
</tbody>
</table>
CHAPTER 3
LOCAL SENSITIVITY ANALYSIS FOR MISSING OUTCOMES AND PREDICTORS

In Section 3.1, we derive expressions for ISNI and related statistics in the setting of missing data in outcomes and predictors, and describe an approach to interpretation. In Section 3.2, we derive the equations in the context of conditional logistic regression. In Section 3.3, we elucidate the index in a simple simulation study via artificial deleting. In Section 3.4, we illustrate the index in two real-data applications involving the normal linear model and conditional logistic regression.

3.1. Methodology

3.1.1. ISNI

The data consist of independently and identically distributed copies of \((Y_i, X_i, Z_i), i = 1, \ldots, N\), where \(Y_i\) is the outcome, \(X_i\) is a predictor that is subject to missingness, and \(Z_i\) is a vector of predictors that are not subject to missingness. \(G_i\) and \(H_i\) are indicators of whether \(Y_i\) and \(X_i\), respectively, are observed: \(G_i = 1(0)\) if \(Y_i\) is observed (missing); \(H_i = 1(0)\) if \(X_i\) is observed (missing). We can readily generalize \(G_i\) and \(H_i\) to vectors for multivariate \(Y_i\) and \(X_i\).
Denote the joint distribution for \((G_i, H_i, X_i, Y_i|Z_i)\),

\[
f_{G_i|H_i, X_i, Y_i|Z_i}(g_i, h_i, x_i, y_i, z_i) = f_{G_i|H_i, X_i, Y_i|Z_i}(g_i|h_i, x_i, y_i, z_i)f_{H_i|X_i, Y_i, Z_i}(h_i|x_i, y_i, z_i)
\]

\[
f_{\theta|X_i, Z_i}(y_i|x_i, z_i)f_{\beta|Z_i}(x_i|z_i).
\]

To simplify notation, we henceforth replace the symbols \(f_{G_i|H_i, X_i, Y_i|Z_i} \), \(f_{H_i|X_i, Y_i, Z_i} \), \(f_{Y_i|X_i, Z_i} \), \(f_{\beta|Z_i} \) by \(a_\xi, b_\gamma, c_\theta, d_\beta \) respectively. The log likelihood is then

\[
l = \sum_{i=1}^{N} \left\{ h_ig_i \left[ \ln a_\xi(g_i|h_i, x_i, y_i, z_i) + \ln b_\gamma(h_i|x_i, y_i, z_i) + \ln c_\theta(y_i|x_i, z_i) + \ln d_\beta(x_i|z_i) \right] \\
+ h_i(1-g_i) \ln \int a_\xi(g_i|h_i, x_i, u, z_i)b_\gamma(h_i|x_i, u, z_i)c_\theta(u|x_i, z_i)d_\beta(x_i|z_i)du \\
+ (1-h_i)g_i \ln \int a_\xi(g_i|h_i, v, y_i, z_i)b_\gamma(h_i|v, y_i, z_i)c_\theta(y_i|v, z_i)d_\beta(v|z_i)dv \\
+ (1-h_i)(1-g_i) \ln \int a_\xi(g_i|h_i, v, u, z_i)b_\gamma(h_i|v, u, z_i)c_\theta(u|v, z_i)d_\beta(v|z_i)dudv \right\}.
\]

We denote the probabilities that \(Y_i \) and \(X_i \) are observed as, respectively,

\[
a_\xi(1|h_i, x_i, y_i, z_i) = q(\xi_0 + \xi_1x_i + \xi_2y_i + \xi_3z_i + \xi_4h_i),
\]

\[
b_\gamma(1|x_i, y_i, z_i) = r(\gamma_0 + \gamma_1x_i + \gamma_2y_i + \gamma_3z_i),
\]

where \(q \) and \(r \) stand for link functions. The primary parameter of interest is \(\theta \), which indexes the conditional distribution of the outcome \(Y \) given the predictors \(X \) and \(Z \). The remaining parameters are nuisance parameters: \(\beta \) governs the distribution of \(X \) given \(Z \); \(\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3) \) governs the missingness mechanism of \(X \) given \(X, Y, \) and \(Z \); and \(\xi = (\xi_0, \xi_1, \xi_2, \xi_3, \xi_4) \) governs the missingness mechanism of \(Y \) given \(H, X, Y, \) and \(Z \). The nonignorability parameters are \(\nu = (\xi_1, \xi_2, \gamma_1, \gamma_2)^T \), in the sense that if \(\nu = 0 \), then the missingness mechanisms are missing at random (MAR). We moreover denote \(\xi' = (\xi_0, \xi_3, \xi_4) \) and \(\gamma' = (\gamma_0, \gamma_3) \) as the subsets of parameters of the missingness mechanisms that do not affect ignorability. We denote estimates of the primary and nuisance parameters, es-
imated by maximum likelihood (MLE) when positing the nonignorability parameters $\nu$, as $(\hat{\theta}(\nu), \hat{\beta}(\nu), \hat{\xi}'(\nu), \hat{\gamma}'(\nu))$.

Troxel et al [58] introduced the *index of local sensitivity to nonignorability (ISNI)* as the basis of an analysis of sensitivity to nonignorability. Their idea is to assess the variability of the MLE of $\theta$ as a function of the nonignorability parameter in the vicinity of the MAR model. We extend their model, which assumes missingness only in the outcome $Y$, to the situation where both outcomes and predictors can be missing.

We begin by taking a first-order Taylor expansion of $(\hat{\theta}(\nu), \hat{\beta}(\nu), \hat{\xi}'(\nu), \hat{\gamma}'(\nu))$ at $\nu = 0$:

$$
\begin{pmatrix}
\hat{\theta}(\nu) \\
\hat{\beta}(\nu) \\
\hat{\xi}'(\nu) \\
\hat{\gamma}'(\nu)
\end{pmatrix}
\approx
\begin{pmatrix}
\hat{\theta}(0) \\
\hat{\beta}(0) \\
\hat{\xi}'(0) \\
\hat{\gamma}'(0)
\end{pmatrix} +
\frac{\partial(\hat{\theta}(\nu), \hat{\beta}(\nu), \hat{\xi}'(\nu), \hat{\gamma}'(\nu))}{\partial \nu}^T \bigg|_{\nu=0} \cdot \nu.
$$

Following [58], we define the index of local sensitivity to nonignorability as

$$
\text{ISNI} = \frac{\partial(\hat{\theta}(\nu), \hat{\beta}(\nu), \hat{\xi}'(\nu), \hat{\gamma}'(\nu))}{\partial \nu} \bigg|_{\nu=0}.
$$

By the implicit function theorem, a general formula for ISNI is

$$
\begin{pmatrix}
\nabla^2 l_{\theta\theta} & \nabla^2 l_{\theta\beta} & \nabla^2 l_{\theta\xi} & \nabla^2 l_{\theta\gamma} \\
\nabla^2 l_{\beta\theta} & \nabla^2 l_{\beta\beta} & \nabla^2 l_{\beta\xi} & \nabla^2 l_{\beta\gamma} \\
\nabla^2 l_{\xi'\theta} & \nabla^2 l_{\xi'\beta} & \nabla^2 l_{\xi'\xi} & \nabla^2 l_{\xi'\gamma} \\
\nabla^2 l_{\gamma'\theta} & \nabla^2 l_{\gamma'\beta} & \nabla^2 l_{\gamma'\xi} & \nabla^2 l_{\gamma'\gamma}
\end{pmatrix}^{-1} 
\begin{pmatrix}
\nabla^2 l_{\theta\nu} \\
\nabla^2 l_{\beta\nu} \\
\nabla^2 l_{\xi'\nu} \\
\nabla^2 l_{\gamma'\nu}
\end{pmatrix} \bigg|_{\nu=0}.
$$
where $\nabla^2 l_{\theta\theta}$ is the second derivative of the log likelihood function with respect to $\theta$ under ignorable model, and other second derivatives follow similarly. We recognize the first factor in ISNI as the variance-covariance matrix of $(\hat{\theta}, \hat{\beta}, \hat{\xi}', \hat{\gamma}')$ under MAR, and the second factor as a measure of the orthogonality of $(\theta, \beta, \xi', \gamma')$ and $\nu$. The assumption of parameter distinctness, i.e., that there are no a priori ties between $(\theta, \beta)$ and $(\xi, \gamma)$, implies that $(\nabla^2 l_{\theta\xi}', \nabla^2 l_{\theta\gamma}', \nabla^2 l_{\beta\xi}', \nabla^2 l_{\beta\gamma}') = 0$ under MAR. We present detailed formulas for calculating ISNI in Appendix B.1.

3.1.2. Interpretation of ISNI

ISNI measures the degree of local sensitivity to nonignorability in the vicinity of the ignorable model. Although ISNI is typically straightforward to compute, it is not invariant to such factors as the scale of measurement of continuous predictors. Therefore we propose a more flexible and interpretable index that evaluates the minimum degree of nonignorability required to cause a maximum negligible distortion in estimates of parameters of interest [7, 10, 33, 41, 50, 59, 66]. Previous works have denoted such a measure as the $c$ index [58].

Assume the nonignorability parameters $\nu$ is $p$-dimensional and each nonignorability parameter links one variable with missing to one missingness indicator. For example, in Section 3.1.1, $\nu$ is a 4-dimensional vector, $(\xi_1, \xi_2, \gamma_1, \gamma_2)^T$, linking $(x_i, y_i, x_i, y_i)^T$ to these missingness indicators in the nonignorable model. We denote the vector $(x_i, y_i, x_i, y_i)^T$ as the set of corresponding variables for $\nu = (\xi_1, \xi_2, \gamma_1, \gamma_2)^T$ in the missingness mechanisms. The vector of corresponding variables has the same dimension as the nonignorability parameter $\nu$.

The primary parameter of interest is $\theta$. We denote

$$\text{ISNI}(\hat{\theta}) = \left(\text{ISNI}_1(\hat{\theta}), \ldots, \text{ISNI}_i(\hat{\theta}), \ldots, \text{ISNI}_p(\hat{\theta})\right),$$

28
where \( ISNI_i(\hat{\theta}) \) is the first derivative of \( \hat{\theta}(\nu) \) with respect to the \( i \)-th nonignorability parameter evaluated at \( \nu = 0 \). If the \( i \)-th element in the corresponding variables is continuous, \( ISNI_i(\hat{\theta}) \) will be scale-dependent [58]. Denote \( \sigma = (\sigma_1, \ldots, \sigma_p) \) for the \( p \) corresponding variables, where \( \sigma_i \) is the standard deviation of the corresponding variable if it is continuous, or 1 if it is discrete. The standardized ISNI is defined as

\[
SISNI(\hat{\theta}) = (SISNI_1(\hat{\theta}), \ldots, SISNI_i(\hat{\theta}), \ldots, SISNI_p(\hat{\theta})) = \left(\frac{ISNI_1(\hat{\theta})}{\sigma_1}, \ldots, \frac{ISNI_i(\hat{\theta})}{\sigma_i}, \ldots, \frac{ISNI_p(\hat{\theta})}{\sigma_p}\right).
\]

We define the minimum nonignorability (MinNI) to be the minimum degree of nonignorability that causes a maximum negligible distortion of \( \hat{\theta} \). As a default, we set the maximum negligible distortion to be the standard error (SE) of \( \hat{\theta} \) under the MAR model. Xie and Heitjan [63] proposed an extended ISNI in \( L^2 \) space by Hölder’s inequality to approximately measure the maximal sensitivity, and we transform it to minimum nonignorability as follows:

\[
\text{MinNI} = \frac{\text{SE}(\hat{\theta})}{\|SISNI\|_2},
\]

where

\[
\|SISNI\|_2 = (SISNI_1(\hat{\theta})^2 + \cdots + SISNI_i(\hat{\theta})^2 + \cdots + SISNI_p(\hat{\theta})^2)^{\frac{1}{2}}.
\]

Algebraically, MinNI is approximately the radius of the smallest ball, centered at the MAR model, needed to produce a 1-SE change of \( \hat{\theta} \). If MinNI is small, then the minimum nonignorability needed to distort \( \hat{\theta} \) is plausible. That is, even modest nonignorability leads to sensitive estimates of parameters. If MinNI is large, only extreme nonignorability results in sensitivity.

Troxel et al [58] suggested a cutoff value 1 for \( c \) index, indicating that the minimal nonignorability to cause 1-SE displacement in \( \hat{\theta} \) is that one unit change in outcome is associated with an odds ratio of 2.7 in the observation of probability. Similarly, we use
a cutoff value of 1 for MinNI, indicating that the minimum radius of a $p$-ball where the nonignorability parameters lie, needed to induce 1-SE distortion of $\hat{\theta}$ is 1. That is, if MinNI < 1, the sensitivity to nonignorable missing should be a serious concern.

As indicated above, Troxel et al [58] proposed the scale-independent sensitivity transformation or $c$ value, which is an one-dimensional version of MinNI. Xie and Heitjan [63] proposed an index, SET, that is a two-dimensional version of MinNI that measures sensitivity to nonignorable treatment crossover in a randomized trial, where the crossover mechanism can differ by treatment arm. Chen [7] has proposed MinNI measures for the situation where missingness results from unmeasured confounders between the missingness indicator and the outcome. Although the unmeasured confounding specification in [7] differs from the selection model, the interpretations of sensitivity values are in the spirit of the proposal of Cornfield [10]. Our proposed MinNI includes the $c$ value and the SET as special cases.

3.2. Conditional Logistic Model

3.2.1. Conditional likelihood

We apply the ISNI analysis to conditional logistic regression in matched case-control studies to assess the degree of local sensitivity when the predictors can have missing observations. The notation is the same as in Section 3.1.1 except that the outcomes are completely observed and the matched strata are defined by another set of completely observed variables $W$. Suppose we have $J$ strata, with stratum $j$ containing 1 case and $M_j$ controls. Subject $i$ in stratum $J$ has data $(h_{ij}, x_{ij}, y_{ij}, z_{ij}, w_{ij})$ for $i = 0, 1, \ldots, M_j$ and $j = 1, \ldots, J$. We denote subject $i = 0$ in each stratum to be the case. The total number of observations is still $N$. 
The joint distribution for \((Y_i, X_i, H_i|Z_i, W_i)\) is,

\[
f_{Y_i,X_i,H_i|Z_i,W_i}(y_i, x_i, h_i|z_i, w_i) = f_{H_i|X_i,Y_i,Z_i,W_i}(h_i|x_i, y_i, z_i, w_i) f_{Y_i,X_i,Z_i,W_i}(y_i|x_i, z_i, w_i) f_{X_i|Z_i,W_i}(x_i|z_i, w_i).
\]

Denote

\[
f_{H_i|X_i,Y_i,Z_i,W_i}(1|x_i, y_i, z_i, w_i) = r(1|x_i, y_i, z_i, w_i).
\]

Without loss of generality, assume \(X\) is discrete with a finite number of levels. We modify the parameterization for \(f_{Y_i|X_i,Z_i,W_i}(y_i|x_i, z_i, w_i)\) as in [52]. Define the odds of \(Y\) conditional on \(X, Z, W\) and the distribution of \(X\) conditional on \(Z, W\) in the control arm to be, respectively,

\[
\eta(x, z, w) = \frac{\Pr[Y = 1|X = x, Z = z, W = w]}{\Pr[Y = 0|X = x, Z = z, W = w]},
\]

\[
\pi(x|z, w) = \Pr[X = x|Y = 0, Z = z, W = w].
\]

After we specify models for \(\eta\) and \(\pi\), the other two functions are determined:

\[
\tilde{\eta}(z, w) = \frac{\Pr[Y = 1|Z = z, W = w]}{\Pr[Y = 0|Z = z, W = w]} = \sum_v \eta(v, z, w)\pi(v|z, w);
\]

\[
\rho(x|z, w) = \Pr[X = x|Y = 1, Z = z, W = w] = \frac{\pi(x|z, w)\eta(x, z, w)}{\tilde{\eta}(z, w)}.
\]

The conditional likelihood for the 1:\(M_j\) matched case-control study is

\[
L = \left\{ \prod_{j=1}^{J} \sum_{k=0}^{M_j} \tilde{\eta}(z_{0j}, w_{0j}) \right\}^{h_i} \prod_{i=1}^{N} \left\{ \pi(x_i|z_i, w_i)(1-y_i)\rho(x_i|z_i, w_i)^{y_i}r(1|x_i, y_i, z_i, w_i) \right\}^{h_i} \left\{ \sum_v \pi(v|z_i, w_i)(1-y_i)\rho(v|z_i, w_i)^{y_i}[1 - r(1|v, y_i, z_i, w_i)] \right\}^{1-h_i}.
\]
3.2.2. Model specification

For brevity, assume that $X$ is binary. Define

$$\eta(x, z, w) = \exp \{\theta_0(w) + V(x, z, w)^T \theta\}$$

and

$$\pi(x | z, w) = \frac{\exp \{xU(z, w)\beta\}}{1 + \exp \{U(z, w)\beta\}}.$$

Thus, the score functions for $\theta$ and $\beta$ under MAR are,

$$\sum_{i=1}^N \left\{ V(x_i, z_i, w_i) y_i h_i + \hat{V}(z_i, w_i) [y_i (1 - h_i) - Y_c(z_i, w_i)] \right\} = 0, \quad (3.1)$$

$$\sum_{i=1}^N \left\{ y_i - Y_c(z_i, w_i) \right\} \hat{U}(z_i, w_i) + h_i \tilde{U}(x_i, z_i, w_i) - y_i h_i \tilde{U}(z_i, w_i) = 0, \quad (3.2)$$

where

$$\tilde{U}(x_i, z_i, w_i) = (-1)^{1-x_i} (1 - \pi(x_i | z_i, w_i)) U(z_i, w_i),$$

$$\hat{V}(z_i, w_i) = \sum_v V(v, z_i, w_i) \rho(v | z_i, w_i),$$

$$Y_c(z_i, w_i) = \frac{\tilde{\eta}(z_i, w_i)}{\sum_{k=0}^{M_{j(i)}} \tilde{\eta}(z_{k,j(i)}, w_{k,j(i)})}$$

with $j(i)$ meaning the $j$-th strata where the $i$-th observation belongs to, and

$$\hat{U}(z_i, w_i) = \sum_v (-1)^{1-v} \rho(v | z_i, w_i) (1 - \pi(v | z_i, w_i)) U(z_i, w_i).$$

We can solve the score equations (3.1) and (3.2) simultaneously through quasi-Newton algorithms with numerical Hessian [6, 16, 18, 53]; this enables us to compute the first fac-
tor of ISNI. Assuming $r$ is a logistic link, the terms in the second factor of ISNI are

\[
\frac{\partial^2 l}{\partial \theta \partial \gamma_1} \bigg|_{\gamma_1=0} = \sum_{i=1}^{N} -(1 - h_i) r_i \left[ y_i \sum_v v \rho(v | z_i, w_i) (V(v, z_i, w_i) - \hat{V}(z_i, w_i)) \right],
\]

\[
\frac{\partial^2 l}{\partial \beta \partial \gamma_1} \bigg|_{\gamma_1=0} = \sum_{i=1}^{N} -(1 - h_i) r_i \left\{ y_i \sum_v v \rho(v | z_i, w_i) \left[ \tilde{U}(v, z_i, w_i) - \hat{U}(z_i, w_i) \right] + (1 - y_i) \sum_v v \pi(v | z_i, w_i) \tilde{U}(v, z_i, w_i) \right\},
\]

where

\[
\tilde{U}(v, z_i, w_i) = (-1)^{1-v}(1 - \pi(v | z_i, w_i)) U(z_i, w_i),
\]

\[
r_i = \frac{\exp \{ \hat{\gamma}_0 + \hat{\gamma}_2 y_i + \hat{\gamma}_3 z_i + \hat{\gamma}_4 w_i \}}{1 + \exp \{ \hat{\gamma}_0 + \hat{\gamma}_2 y_i + \hat{\gamma}_3 z_i + \hat{\gamma}_4 w_i \}}
\]

under MAR. Then, plug all the estimations into ISNI formula and modify it to MinNI as defined in Section 3.1.2.

### 3.3. Simulated Missing Observations in the Smoking and Mortality Data

We illustrate our proposed ISNI and MinNI by artificially deleting observations from a complete data set. The data are from a smoking and mortality study of English men grouped into 25 occupational categories [42]. There are two variables: The smoking index (the predictor) is the ratio of the average number of cigarettes smoked per day by men in the occupational group to the average number of cigarettes smoked per day by all men; the mortality index (the outcome) is the ratio of the rate of deaths from lung cancer among men in the occupational group to the rate of deaths from lung cancer among all men. The slope in a linear regression is 1.088 with SE 0.221.

To systematically delete observations, we first order the data according to the values of the smoking index and then, delete the smoking index points sequentially by ranks. Similarly, we order the data by the mortality index to perform deleting of the mortality
index. We construct four types of missing patterns: A single point missing on mortality index with a single point missing on smoking index; a single point missing on mortality index with five points missing on smoking index; five points missing on mortality index with a single point missing on smoking index; and five points missing on mortality index with five points missing on smoking index. Assume the smoking index and the mortality index follow a bivariate normal distribution. We present the most and least sensitive cases of each type in Table 3.1.

In the first missingness type, the least sensitive case is the one that omits point 7 of the mortality index and point 21 of the smoking index. This gives a MinNI of 80.148, which says that the minimum radius of a $4$-ball of vectors of nonignorability parameters needed to cause 1 SE change in the slope estimation is 80.148. This suggests the needed minimum nonignorability is implausible and thus the MLE estimation of the slope is insensitive. The most sensitive case omits point 1 of the mortality index and point 2 of the smoking index, giving a MinNI of 1.639.

The least and the most sensitive cases with their MLEs, standard errors and MinNIs under the second, third and fourth types missingness types appear in rows three to eighth of Table 3.1. When we move from the first type to the fourth type, the MinNIs for the least sensitive case are decreasing (i.e. the estimates getting more sensitive) as the proportion of missing values increasing. Generally, when data are missing toward the middle of the range of smoking status, sensitivity is modest, because the missing points have low leverage and cannot readily influence estimation of the slope. Conversely, when missing points are at the edge of the range of smoking status, sensitivity can be substantial, because these are points of high influence [58].
3.4. Real-Data Examples

3.4.1. The New York School Choice Experiment

The New York School Choice Experiment, conducted in 1997, sought to estimate the effect of vouchers to attend private school on the academic performance of children from low-income families in New York City [4, 27]. The data consist of 525 selected children and 525 matched controls with a list of predictor variables spanning educational, demographic, and socioeconomic indicators and baseline academic performance. The outcome variables were reading and math scores in the school year after the randomization. See more details about the data in [27].

To illustrate our method, we will take the math score to be the outcome variable, with all the other variables as predictors. The model is a multivariate normal linear regression. We conducted a complete-case analysis using elastic net regularization to identify a small subset of strong predictors. We used default settings of the `cv.glmnet` function in R package `glmnet` [17]. The analysis identified as important predictors the grade level and the pre-test math score. We also included the randomization indicator, as the main purpose of the study was to evaluate its effect. The missingness patterns for the full data set (including both complete and incomplete cases) appear in Table 3.2.

Assume the distribution for the pre-test and the post-test math scores conditional on grade level and randomization follows bivariate normal. We estimated coefficients of the regression of post-test math score on pre-test math score, three indicators of grade level, and randomization indicator. First we conducted a complete-case analysis, as shown in Column 2 of Table 3.3. Next we computed the maximum likelihood estimate of the coefficients using the full data set, integrating the density over missing observations to obtain the likelihood function under MAR; see Column 3 of Table 3.3. Corresponding MinNI values for the regression coefficients appear in Column 4.
All MinNIs for these predictors are greater than 1, but the one for randomization is close to 1. The coefficient of randomization measures the encouragement effect of being offered a voucher on math scores. Thus the estimated effect from complete cases and under MAR, which are of borderline significance, are potentially sensitive to nonignorability. The SISNI vector for lottery status is \((-0.051, -0.902, 0.065, 0.103)^T\). The largest magnitude of the elements in this SISNI vector is 0.902, corresponding to the missing post-test math score in the missingness mechanism for the post-test math score. The main contribution to the overall local sensitivity on the parameter estimation of lottery status is from the missing post-test math score in the missingness mechanism for the post-test math score.

3.4.2. The Los Angeles Endometrial Cancer Case Control Study

This was a 1:4 matched case-control study that investigated the effect of various risk factors on endometrial cancer, conducted among residents of the Leisure World retirement community. Investigators matched 63 cases to 4 controls each by date of birth, marital status, and residence [39, 52]. The explanatory variables of interest are GALL (history of gall bladder disease), OB (obesity), and EST (history of use of estrogen therapy). Only OB has missing observations, with 50 (16%) of the values unobserved. Because no stratification variables are available in the data set, we assumed that \(\pi(x|z, w)\) did not depend on \(w\).

Table 3.4 displays the complete-case and maximum likelihood estimates of the regression coefficients, together with ISNI and MinNI values. All the MinNIs are greater than 1. But in this case several of the predictors are relatively sensitive to potential nonignorability compared with other predictors, including main effects and interactions involving OB but also other main effect EST. In particular, inferences regarding the effects of OB, EST, \(OB \times GALL\) and \(OB \times EST\) in this data should be regarded with caution.
Table 3.1: Sensitivity analysis for the slope in the smoking data, with artificially deleted data

<table>
<thead>
<tr>
<th>Missing Mortality Ranks</th>
<th>Missing Smoking Ranks</th>
<th>MLE</th>
<th>SE</th>
<th>MinNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>21</td>
<td>1.026</td>
<td>0.211</td>
<td>80.148</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.815</td>
<td>0.261</td>
<td>1.639</td>
</tr>
<tr>
<td>16</td>
<td>14-18</td>
<td>1.120</td>
<td>0.214</td>
<td>349.218</td>
</tr>
<tr>
<td>5</td>
<td>1-5</td>
<td>0.433</td>
<td>0.349</td>
<td>0.472</td>
</tr>
<tr>
<td>5-9</td>
<td>12</td>
<td>1.112</td>
<td>0.204</td>
<td>371.535</td>
</tr>
<tr>
<td>1-5</td>
<td>5</td>
<td>0.505</td>
<td>0.343</td>
<td>0.415</td>
</tr>
<tr>
<td>9-13</td>
<td>6-10</td>
<td>1.277</td>
<td>0.251</td>
<td>64.804</td>
</tr>
<tr>
<td>5-9</td>
<td>1-5</td>
<td>0.261</td>
<td>0.268</td>
<td>0.303</td>
</tr>
</tbody>
</table>

Table 3.2: Missingness Patterns in the New York School Choice Experiment Data

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Number of Observations</th>
<th>Missing Proportion(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>Post-test Math Score</td>
<td>817</td>
</tr>
<tr>
<td>Predictors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test Math Score</td>
<td>961</td>
<td>8.5</td>
</tr>
<tr>
<td>Grade Level</td>
<td>1050</td>
<td>0</td>
</tr>
<tr>
<td>Randomization</td>
<td>1050</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.3: Sensitivity Analysis for the New York School Choice Experiment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Complete case (SE)</th>
<th>MLE under MAR (SE)</th>
<th>MinNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test Math Score</td>
<td>0.418 (0.033)</td>
<td>0.416 (0.033)</td>
<td>5.648</td>
</tr>
<tr>
<td>Grade Level(2)</td>
<td>-3.248 (1.657)</td>
<td>-3.195 (1.636)</td>
<td>5.491</td>
</tr>
<tr>
<td>Grade Level(3)</td>
<td>5.964 (1.638)</td>
<td>6.322 (1.624)</td>
<td>2.190</td>
</tr>
<tr>
<td>Grade Level(4)</td>
<td>1.355 (1.873)</td>
<td>1.791 (1.822)</td>
<td>4.436</td>
</tr>
<tr>
<td>Randomization</td>
<td>1.922 (1.216)</td>
<td>2.157 (1.197)</td>
<td>1.313</td>
</tr>
</tbody>
</table>
Table 3.4: Sensitivity Analysis for the LA Endometrial Cancer Study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Complete case MLE(SE)</th>
<th>Conditional Logistic MLE under MAR(SE)</th>
<th>ISNI</th>
<th>MinNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>OB</td>
<td>1.460(1.381)</td>
<td>1.291(1.584)</td>
<td>0.521</td>
<td>3.039</td>
</tr>
<tr>
<td>GALL</td>
<td>3.256(1.282)</td>
<td>2.897(1.103)</td>
<td>0.202</td>
<td>5.463</td>
</tr>
<tr>
<td>EST</td>
<td>3.531(1.395)</td>
<td>3.200(1.451)</td>
<td>0.454</td>
<td>3.195</td>
</tr>
<tr>
<td>OB×GALL</td>
<td>-0.146(0.914)</td>
<td>-0.087(0.665)</td>
<td>-0.180</td>
<td>3.686</td>
</tr>
<tr>
<td>OB×EST</td>
<td>-1.116(1.389)</td>
<td>-0.762(1.603)</td>
<td>-0.507</td>
<td>3.162</td>
</tr>
<tr>
<td>GALL×EST</td>
<td>-2.270(1.168)</td>
<td>-2.020(0.936)</td>
<td>-0.075</td>
<td>12.482</td>
</tr>
</tbody>
</table>
Chapter 4
Sensitivity of Estimands in Clinical Trials with Noncompliance

In Section 4.1, we supplement the basic Rubin causal model (RCM) for a clinical trial to reflect potential association between compliance and outcome. Within this framework, we illustrate the effect of compliance on the ITT and CACE estimands. We moreover demonstrate a condition, which we denote the “sixth assumption”, that is sufficient to render CACE robust to such variation. This assumption is plausible though unverifiable in any single study, but with our model one can readily conduct analyses to illustrate sensitivity to its violation. In Sections 4.2 and 4.3, we illuminate such analyses through simple simulation studies and a trial of vitamin A supplementation in children.

4.1. Clinical Trial Estimands

4.1.1. The RCM: Notation

Consider a population with \( N \) experimental units, \( i = 1, \ldots, N \), whom we will randomize between two study arms. Denote the \( N \)-vector of randomization assignments \( Z = (Z_1, \ldots, Z_N) \), where \( Z_i = 1(0) \) indicates assignment of subject \( i \) to the experimental (control) arm. In many clinical trials, subjects can exercise some control over the treatment they receive, in which case the treatment received may not match the treatment assigned. Thus, the assignment vector \( Z \) gives rise to a further \( N \)-vector of actual treatments received \( D = D(Z) \) with \( i \)-th element \( D_i(Z) \). Here, \( D_i(Z) = 1(0) \) indicates that, for treatment assignment vector \( Z \), unit \( i \) receives the experimental (control) treatment. The
outcome is denoted as $Y = Y(Z, D)$ with $Y_i(Z, D(Z))$ indicating the $i$-th outcome value given the assignment $Z$ and the treatment received $D(Z)$.

Note that both the treatment received and the outcome are potential outcomes, in that there are as many potential $N$-vectors $D(Z)$ and $Y(Z, D(Z))$ as there are values of the randomization vector $Z$. If, as is often the case, there is no interference between units, it is possible to make the stable unit treatment value assumption (SUTVA) (see Section 4.1.2), which asserts that $D_i(Z) = D_i(Z_i)$ and $Y_i(Z, D(Z)) = Y_i(Z_i, D_i(Z_i))$. This greatly simplifies the model, because we need to consider only two potential values of the treatment received — $D_i = (D_i(0), D_i(1))$ — and four potential values of the outcome — $Y_i = (Y_i(0, 0), Y_i(0, 1), Y_i(1, 0), Y_i(1, 1))$. Moreover, among the four potential outcomes in $Y_i$ we need consider only the two that can actually arise: $Y_i(0, D_i(0))$ and $Y_i(1, D_i(1))$.

Considering the various patterns of $D_i(0)$ and $D_i(1)$ leads to four principal strata of compliance behaviors; we denote this variable $T_i(D_i)$:

$$
T_i = \begin{cases} 
n \text{ ("never-taker"),} & \text{if } (D_i(0), D_i(1)) = (0, 0); 

c \text{ ("complier"),} & \text{if } (D_i(0), D_i(1)) = (0, 1); 

a \text{ ("always-taker"),} & \text{if } (D_i(0), D_i(1)) = (1, 1); 

d \text{ ("defier"),} & \text{if } (D_i(0), D_i(1)) = (1, 0). 
\end{cases}
$$

Subject $i$ has a vector of data $(Z_i, D_i(0), D_i(1), Y_i(0, D_i(0)), Y_i(1, D_i(1)))$, of which only the three elements $(Z_i, D_i(Z_i), Y_i(Z_i, D_i(Z_i)))$ are observable. We define $D^{\text{obs}}$ to be the $N$-vector of treatment taken $D^{\text{obs}} = D(Z)$, and the realized outcome of interest as $Y^{\text{obs}} = Y(Z, D^{\text{obs}})$. In the basic RCM, the sole random element is $Z$; the other variables are fixed but possibly unknown constants, analogous to the role of outcome variables in design-based sampling theory [2, 31, 32].
4.1.2. Trial estimands viewed in light of the RCM

To define causal estimands in the finite population, we introduce a notation to denote
the average sign, \( E_{\text{fp}} \), over the fixed \( N \)-subject population. For example, define the aver-
age causal effect of randomization on treatment received as,

\[
E_{\text{fp}}[D_i(1) - D_i(0)] = \frac{1}{N} \sum_{i=1}^{N} [D_i(1) - D_i(0)].
\]

We further denote this quantity as the *intention-to-treat (ITT)* estimand for treatment re-
ceived, \( \text{ITT}^{(D)} \). Similarly, the ITT estimand for outcome is,

\[
\text{ITT}^{(Y)} = E_{\text{fp}}[Y_i(1, D_i(1)) - Y_i(0, D_i(0))],
\]

which is the average causal effect on \( Y \) of random assignment, averaged across all com-
pliance classes.

Angrist et al [2] identified five key assumptions that permit identification of an alterna-
tive informative causal estimand, the *complier average causal effect (CACE)*:

1. The *stable unit treatment value assumption (SUTVA)* asserts that there is no inter-
ference among subjects and that the treatment is unique and stable (see Section
   4.1.1 above).
2. *Random assignment* states that the treatment assignment \( Z \) is randomized.
3. The *exclusion-restriction* indicates that any effect of \( Z \) on \( Y \) depends only on \( D \); that
   is, \( Y(Z, D(Z)) = Y(D(Z)) \).
4. There is a *non-zero average causal effect of \( Z \) on \( D \);* that is, \( E_{\text{fp}}[D_i(1) - D_i(0)] \neq 0 \).
5. *Monotonicity*, which asserts that there are no defiers; this further implies that \( E_{\text{fp}}[D_i(1) - 
   D_i(0)] \) is the proportion of compliers.
Under Assumptions 1 and 3, we define CACE as

$$\text{CACE} = \mathbb{E}_{i} [Y_i(1) - Y_i(0) | D_i(1) - D_i(0) = 1],$$

which is the average effect on $Y$ of treatment received in the latent subset of subjects who would take the assigned treatment, whatever it is. Under Assumptions 2, 3 and 4, $Z$ is a valid instrumental variable. Then, Assumptions 1, 3, 4 and 5 are sufficient to guarantee that

$$\text{ITT}^{(Y)} = \text{CACE} \times \text{ITT}^{(D)}.$$

Thus, under the five assumptions, CACE equals the ratio of $\text{ITT}^{(Y)}$ and $\text{ITT}^{(D)}$, and is denoted the instrumental variables estimand \cite{1, 2, 15, 31, 32, 48, 56}.

Setting $\hat{\text{ITT}}^{(D)} = \frac{\sum D_{\text{obs}} D_i}{\sum Z_i} - \frac{\sum D_{\text{obs}} (1-Z_i)}{\sum (1-Z_i)}$ to be the observed effect of assignment on treatment received, and $\hat{\text{ITT}}^{(Y)} = \frac{\sum Y_{\text{obs}} Z_i}{\sum Z_i} - \frac{\sum Y_{\text{obs}} (1-Z_i)}{\sum (1-Z_i)}$ to be the observed effect of assignment on outcome, Assumptions 1 and 2 imply that $\hat{\text{ITT}}^{(Y)}$ and $\hat{\text{ITT}}^{(D)}$ are unbiased for $\text{ITT}^{(Y)}$ and $\text{ITT}^{(D)}$, respectively. Together, the five assumptions guarantee that

$$\hat{\text{CACE}} = \frac{\hat{\text{ITT}}^{(Y)}}{\hat{\text{ITT}}^{(D)}}$$

is a consistent estimator of CACE.

4.1.3. The sixth assumption

4.1.3.1. Definition

In the conventional derivation of CACE, compliance behavior is a fixed characteristic of study subjects. Factors that influence compliance in clinical trials may include actual toxicities, concern about toxicities, the severity of the condition being treated, and subject
confidence in the treatment and the physicians who are administering it. Thus, attitudes toward compliance may vary between trials, potentially inducing sensitivity of CACE.

Assume that both compliance class, determined by \((D_i(0), D_i(1))\), and the outcome vector \(Y_i\) are defined for subjects in a super-population from which we will sample for the clinical trial. The ITT estimands and the five assumptions are defined in the same way as in the finite-population case after the random sampling of subjects and random assignment of treatments. We use the notation \(E_{sp}\) to represent averaging over the superpopulation distribution.

Under Assumption 1, define the \textit{total average causal effect} (TACE) as

\[
TACE = E_{sp}[Y_i(1, 1) - Y_i(0, 0)],
\]

which equals \(E_{sp}[Y_i(1) - Y_i(0)]\), denoted as PACE, if Assumption 3 holds. Define the \textit{total complier average causal effect} as

\[
TCACE = E_{sp}[Y_i(1, 1) - Y_i(0, 0) | D_i(1) - D_i(0) = 1],
\]

which equals \(E_{sp}[Y_i(1) - Y_i(0) | D_i(1) - D_i(0) = 1]\), denoted as CACE, again under Assumption 3. When compliance is correlated with the individual causal effect, TCACE will vary as compliance varies. We define two versions of the sixth assumption relating outcome and compliance in the superpopulation:

6(a): Compliance is independent of the individual causal effect of randomization and treatment; i.e., \((D_i(1), D_i(0)) \perp \perp Y_i(1, 1) - Y_i(0, 0)\).

6(b): Compliance is independent of the individual causal effect of treatment under both randomization groups; i.e., \((D_i(1), D_i(0)) \perp \perp Y_i(z_i, 1) - Y_i(z_i, 0), \text{ for } z_i = 0, 1\).

Assumptions 6(a) and 6(b) differ in the interpretations of individual causal effects but are
equivalent when Assumption 3 holds. Under Assumptions 1 and 6(a),

\[ \text{TCACE} = \mathbb{E}_\text{sp}[Y_i(1, 1) - Y_i(0, 0)|D_i(1) - D_i(0) = 1] = \mathbb{E}_\text{sp}[Y_i(1, 1) - Y_i(0, 0)] = \text{TACE}. \tag{4.1} \]

Under Assumptions 1 and 6(b),

\[ \mathbb{E}_\text{sp}[Y_i(z_i, 1) - Y_i(z_i, 0)|D_i(1) - D_i(0) = 1] = \mathbb{E}_\text{sp}[Y_i(z_i, 1) - Y_i(z_i, 0)]. \tag{4.2} \]

If Assumption 3 also holds, the left-hand side of Equation 4.1 or 4.2 equals CACE and the right-hand side equals PACE, indicating that CACE is robust to varying compliance. The first five assumptions guarantee identification of CACE through the instrumental variable \( Z \), whereas Assumption 6(b), henceforth referred to as “the sixth assumption”, guarantees its robustness when the compliance class is considered random. This assumption is rather plausible but appears to be impossible to validate in any single study, as it refers to association between two partial observables.

\subsection*{4.1.3.2. Relationship to other assumptions}

“The sixth assumption” (A6) does not imply unconfoundedness between \( D_i \) and \( Y_i \) because \( D_i(z_i) \) is typically correlated with \( Y(z_i, D_i(z_i)) \) due to unmeasured confounding, but \( Y(z_i, 1) - Y(z_i, 0) \) could be independent of \( D_i(z_i) \). Thus, estimands based on grouping by \( D_i(z_i) \) are biased. A6 implies that any latent compliance subpopulation causal effect is equal to PACE. If monotonicity is violated, CACE is an average causal effect of treatment received in compliers and defiers. If A6 is valid, there is no need of a monotonicity assumption to identify the latent subset of compliers, because the interpretation is the average causal effect of treatment received in the entire population. For simplicity in sensitivity analysis, we continue to posit monotonicity in simulation and real data analysis.

A6 is sufficient to guarantee robustness of CACE when compliance class is random.
To clarify the uniqueness and importance of A6, we further demonstrate its relations and comparisons with other homogeneity conditions. Hernán and Robins [25] summarize two closely related homogeneity conditions:

H1. The effect of treatment received $D$ on outcome $Y$ is constant across individuals [5, 40].

H2. For dichotomous $Z$ and $D$, the average causal effect of $D$ on $Y$ across levels of $Z$ in both the treated and in the untreated are equal [21, 49].

The aims of A6, H1 and H2 are different. The homogeneity conditions from Hernán and Robins [25] are constructed to demonstrate the equality of local average causal effect and PACE under structural equation models [26, 61]. However, we aim to address instability in CACE when the compliance status is regarded as random. One potential way to construe robustness is by imposing equality between CACE and PACE, which shares the same consequences as the homogeneity conditions. A6 is implied by and weaker than H1. A6 has wider scope of generalizability than H2, because H2 requires the assignment and treatment received to be dichotomous, whereas A6 can be readily generalized to any type of variable. For a detailed comparison of structural equation models and the RCM see Angrist et al [2].

4.2. Simulations

4.2.1. A simple model relating compliance to outcome

Suppose that the first five assumptions hold. Assume also that there is a latent, continuous compliance variable, denoted $W = (W_1, \ldots, W_N)$, and that $T_i = n \iff W_i < A_1$, $T_i = c \iff A_1 \leq W_i \leq A_2$, $T_i = a \iff W_i > A_2$ for some constants $A_1 < A_2$ that reflect the trial’s overall level of compliance. We assume moreover that in the superpopulation,
\( W_i \) and \( Y_i(1) - Y_i(0) \) follow the bivariate normal distribution,

\[
\begin{pmatrix}
  W_i \\
  Y_i(1) - Y_i(0)
\end{pmatrix}
\sim 
\mathcal{N}
\begin{pmatrix}
  0 \\
  \tau
\end{pmatrix},
\begin{pmatrix}
  1 & \rho \\
  \rho & 1
\end{pmatrix}.
\]

Thus \( \text{PACE} = \tau \), and \( \rho \) describes the correlation between the individual compliance and the individual causal effect.

Assume for the moment that in addition to there being no defiers there are also no always-takers, as would be the case when the experimental treatment is only available through participation in the trial. We can achieve this by setting \( A_2 = \infty \), which implies \( \Pr[T_i = a] = 0 \) and therefore that there are only compliers and never-takers. We then derive

\[
\text{CACE} = E_{sp}[Y_i(1) - Y_i(0)|W_i \geq A_1] = \tau + \rho \frac{\phi(A_1)}{1 - \Phi(A_1)},
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are the standard normal density and distribution, respectively. If we vary \( A_1 \) to change the proportion of compliers, CACE will also change unless \( \rho = 0 \), implying that \( W_i \) and \( Y_i(1) - Y_i(0) \) are independent under the normal model. This verifies the requirement of the sixth assumption for the robustness of CACE.

To illustrate the relationship between CACE and the proportion of compliers, we generate a population of size \( N = 10,000 \) under parameters

- \( \tau = 2 \),
- \( \rho \in \{ -0.6, -0.3, 0, 0.3, 0.6 \} \),

and let \( A_1 \) vary in \((-2.5, 2.5)\) and \( Y_i(0) \sim N(W_i, 1) \). Figure 4.1 plots values of CACE and ITT\(^{(Y)}\) as functions of the proportion of compliers. If \( \rho = 0 \), the sixth condition is satisfied, and CACE = 2 regardless of the fraction of compliers. For any \( \rho \neq 0 \), CACE varies with the proportion of compliers, equaling PACE only when all subjects comply. ITT\(^{(Y)}\) is generally
less variable as the proportion of compliers approaches 0, but only equals PACE when there is full compliance. Even when there is no association between compliance status and individual causal effects, ITT\(^{(Y)}\) differs from PACE if there is any noncompliance.

4.2.2. Model with always-takers

To allow the possibility of always-takers, we set \(A_2 < \infty\). We then derive

\[
\text{CACE} = E_{\text{sp}}[Y_i(1) - Y_i(0) | A_1 \leq W_i \leq A_2] = \tau + \rho \frac{\phi(A_1) - \phi(A_2)}{\Phi(A_2) - \Phi(A_1)}.
\]

Again, \(\rho = 0\) implies \(\text{CACE} = \tau\) regardless of the fraction of noncompliance. We now generate a population of size \(N = 10,000\) assuming \(\tau = 2\), \(\rho \in \{0, 0.6\}\), and different combinations of \(A_1 < A_2\) for \(A_1 \in (-2.5, 2.5)\) and \(A_2 \in (-2.4, 2.6)\).

We constructed 3D plots to illustrate variation in CACE and ITT\(^{(Y)}\). Seen from (1) and (2) in Figure 4.2, CACE and ITT\(^{(Y)}\) vary with \(A_1\) and \(A_2\) when \(\rho \neq 0\), and they are more unstable as the proportion of compliers approaches zero. When the sixth assumption holds (i.e., \(\rho = 0\) in the normal model), only CACE is robust to a varying proportion of compliers, as shown in panel (3) of Figure 4.2. ITT\(^{(Y)}\) changes with the proportion of compliers, as seen in panel (4) of Figure 4.2. When \(A_1\) is fixed at a small value and \(A_2\) increases, the proportion of compliers increases and CACE and ITT\(^{(Y)}\) approach PACE. If \(A_1\) is large, the proportion of compliers is small irrespective of the value of \(A_2\).

4.3. Illustrative Example

4.3.1. The Vitamin A Supplement Data

We assess the sensitivity of inferences to the sixth assumption using data from a trial
of vitamin A supplementation in Indonesia [55]. The study randomized at the village level; 12,094 infants resided in villages that were randomized to vitamin A supplementation, and 11,588 resided in villages that received no supplementation. In the experimental arm, 9,675 infants complied with their random assignment and 2,419 did not. All of the infants assigned to control received control. Seventy-four children in the control group died, compared to 46 in the vitamin A group. Table 4.1 summarizes results.

From these data we compute the ITT estimates:

\[
\hat{\text{ITT}}^{(D)} = \frac{9,675}{12,094} = 0.80; \quad \hat{\text{ITT}}^{(Y)} = \frac{46}{12,094} - \frac{74}{11,588} = -0.0026.
\]

Thus,

\[
\hat{\text{CACE}} = \frac{\hat{\text{ITT}}^{(Y)}}{\hat{\text{ITT}}^{(D)}} = -0.0032.
\]

4.3.2. A latent variable model for outcome

We cannot directly apply the model of Section 4.2.1 because the outcome is binary. Define \( Y_i(1) \) to be the mortality indicator if assigned to supplement and \( Y_i(0) \) to be the mortality indicator if assigned to control. Following a similar approach to the model for compliance, we assume a latent, continuous \( Y^* \) and constants \( B_1 < B_2 \) such that \( Y_i(1) - Y_i(0) = -1 \iff Y^*_i < B_1, Y_i(1) - Y_i(0) = 0 \iff B_1 \leq Y^*_i \leq B_2, \) and \( Y_i(1) - Y_i(0) = 1 \iff Y^*_i > B_2. \) We assume moreover that for each infant, \( W_i \) and \( Y^*_i \) follow the bivariate normal distribution

\[
\begin{pmatrix} W_i \\ Y^*_i \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right].
\]

Thus \( \text{PACE} = 1 - \Phi(B_2) - \Phi(B_1), \) and \( \rho \) describes the correlation between the compliance and the individual causal effect of vitamin A supplement in the superpopulation. To reflect
that the study did not permit always-takers, we set \( A_2 = \infty \). Under these assumptions,

\[
CACE = E_{sp}[Y_i(1) - Y_i(0)|W_i \geq A_1] \\
= 1 + \frac{Pr_\rho[Y_i^* \leq B_2, W < A_1] + Pr_\rho[Y_i^* < B_1, W < A_1] - \Phi(B_2) - \Phi(B_1)}{1 - \Phi(A_1)}.
\]

Note that the two probability terms depend on \( \rho \).

4.3.3. Sensitivity analysis

To investigate the impact of violation of the sixth assumption, we fix \( \rho, B_1, B_2 \) and allow \( A_1 \) to vary, observing the variation in CACE. Let \( PACE = 1 - \Phi(B_2) - \Phi(B_1) = -0.0032 \), the estimated CACE in the real data, and arbitrarily select a combination of \( B_1 \) and \( B_2 \). Then a set of possible values for the parameters is

- \( B_1 = -2, B_2 = 2.0631; \)
- \( \rho \in \{-0.6, -0.3, 0, 0.3, 0.6\}; \)
- \( A_1 \in \{-4, -1.5, -0.84, 0, 0.5, 1, 3\}. \)

Table 4.2 presents variation in \( ITT(Y) \) and CACE as a function of the proportions of compliers when other parameters are held fixed. When \( \rho = 0 \), \( CACE = PACE = -0.0032 \), but \( ITT(Y) \) varies as the proportion of compliers changes. If \( \rho \neq 0 \), both CACE and \( ITT(Y) \) vary with the proportion of compliers. We can locate the estimates of \( ITT(Y) \) and CACE from the observed real data at cells with \( \rho = 0 \) and \( 1 - \Phi(A_1) = 0.80 \). If we fix \( 1 - \Phi(A_1) = 0.80 \), both \( ITT(Y) \) and CACE are sensitive to changes in \( \rho \), possibly even switching signs. If the proportion of compliers is 1, the two estimands are equal to PACE. When the proportion of compliers approaches 0, \( ITT(Y) \) also approaches 0, but CACE is unstable.
Figure 4.1: CACE and ITT\(^{(Y)}\) as functions of \(\rho\) and the proportion of compliers for fixed \(\tau = 2\).

<table>
<thead>
<tr>
<th>Study arm</th>
<th>Compliance</th>
<th>Children</th>
<th>death</th>
<th>Mortality (per 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>-</td>
<td>11,588</td>
<td>74</td>
<td>6.4</td>
</tr>
<tr>
<td>Experimental</td>
<td>-</td>
<td>12,094</td>
<td>46</td>
<td>3.8</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>9,675</td>
<td>12</td>
<td>1.2</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>2,419</td>
<td>34</td>
<td>14.1</td>
</tr>
</tbody>
</table>
Figure 4.2: CACE and ITT\(^{(Y)}\) as functions of \(A_1\) and \(A_2\) for fixed \(\tau = 2\). (1) CACE with \(\rho = 0.6\); (2) ITT\(^{(Y)}\) with \(\rho = 0.6\); (3) CACE with \(\rho = 0\); (4) ITT\(^{(Y)}\) with \(\rho = 0\).
Table 4.2: ITT\(^{(Y)}\) and CACE as a function of the proportion of compliers, \(1 - \Phi(A_1)\).

<table>
<thead>
<tr>
<th>(\rho)</th>
<th>(1 - \Phi(A_1))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>-0.6</td>
<td>ITT(^{(Y)})</td>
</tr>
<tr>
<td></td>
<td>CACE</td>
</tr>
<tr>
<td>-0.3</td>
<td>ITT(^{(Y)})</td>
</tr>
<tr>
<td></td>
<td>CACE</td>
</tr>
<tr>
<td>0.0</td>
<td>ITT(^{(Y)})</td>
</tr>
<tr>
<td></td>
<td>CACE</td>
</tr>
<tr>
<td>0.3</td>
<td>ITT(^{(Y)})</td>
</tr>
<tr>
<td></td>
<td>CACE</td>
</tr>
<tr>
<td>0.6</td>
<td>ITT(^{(Y)})</td>
</tr>
<tr>
<td></td>
<td>CACE</td>
</tr>
</tbody>
</table>
5.1. Sensitivity analysis via unmeasured confounding

We have used a model that frames nonignorability as a consequence of unobserved confounding to devise a simple, general paradigm for sensitivity analysis in frequentist inference of incomplete data. The interpretation of nonignorability in our model mirrors the methods for analysis of sensitivity to nonignorable confounding of Cornfield et al [10], Rosenbaum and Rubin [50], and Ding and VanderWeele [12].

Our sensitivity analysis involves identifying the minimum degree of nonignorability that causes a designated discrepancy in some comparison of \( f^Y(y) \) and \( f^{Y|G}(y|g) \). In the analyses we demonstrate here, we assume that the outcomes \( Y \) represent an i.i.d. sample from some distribution, but that is not essential to the method. Moreover we have derived equations for comparisons of means and variances, but it should be possible to extend the analysis to other functionals. To be conservative, we make use of the partial identification region on the ratio scale to propose the MinNI under the most sensitive case so as to summarize the degree of nonignorability.

When there is a continuous measured covariate \( X \), it would be natural to apply parametric sensitivity models that require specification of a distributional form for \( X \) [33, 41]. If the dimension of \( X \) is large, the nonparametric analysis of MinNI described in Section 2.5.3 will be difficult because of the sparsity of data (i.e. the curse of dimensionality).
Thus, other nonignorable models based on selection specification, such as ISNI analysis or the propensity score matching approach might be much simpler [38, 58, 64–66]. Under parametric sensitivity analysis, one possible extension should consider sensitivity to parametric model misspecification, for instance in link functions relating $Y$ and $G$ to $U$ [20].

Our model differs from many prior developments in this area in referring to frequentist rather than likelihood/Bayesian estimation [22–24]. In principle, we could construct a similar analysis replacing the selection specification with the confounding specification as the incompleteness mechanism in a model-based analysis, and evaluating, say, the minimum nonignorability needed to deflect maximum likelihood estimates by a designated amount. This would render our approach comparable to the ISNI sensitivity analysis.

5.2. Local sensitivity analysis for missing outcomes and predictors

We proposed a generalized index, denoted MinNI, that measures the minimum non-ignorability that gives rise to non-negligible distortion of the maximum likelihood estimate of a parameter of interest when some data are missing on predictors or outcomes. MinNI generalizes other similar indices (known as $c$ or SET) that cover only the situation where the outcome is missing [58, 63, 64]. The simulation study in Section 3.3 and the real-data examples in Section 3.4.1 illustrate application and interpretation of the analysis.

Shi et al [54] proposed a general framework for local sensitivity analysis in generalized linear models with missing covariates. They aimed to identify influential points and test model misspecification. By contrast, we seek to generalize the ISNI analysis of [58], extending it to the setting where both outcomes and predictors can be missing and improving interpretability of the sensitivity index. The idea in our work is to compute the estimate of the parameter of interest under a range of nonignorability assumptions, mapping out the set of assumptions for which change from the MAR estimate is negligible. If
this set is large, then the estimates are insensitive. If it is small, then nonignorability is potentially of concern.

The proposed index, MinNI can be applied without difficulty to generalized linear models under both cohort and case-control sampling. In survival models, we would need to embed the analysis in the general coarse data model [24]. Zhang and Heitjan [64] extended ISNI to the case with nonignorable censoring of an outcome variable. Similarly, we could extend our index to survival models with censoring outcomes and missing predictors.

We envision using the MinNI as a screening index that gives a first-order approximation to the degree of sensitivity. The quality of the approximation is a topic for further study. When the number of nonignorability parameters is too large, our analytic derivations of likelihood will become infeasible. A general, algorithmic approach that involves computing $\hat{\theta}(\nu)$ over a broad range of $\nu$ values and mapping the set of nonignorability parameters that lead to sensitivity may be necessary in models that are more complex or have many parameters.

5.3. Sensitivity of estimands in clinical trials with noncompliance

We demonstrate that a sufficient condition for CACE to be independent of compliance level is that individual compliance and treatment effects are independent. Thus if, as seems plausible in many applications, subjects with larger individual treatment effects are also more likely to comply, CACE will vary with the fraction of compliers. In contrast, ITT\(^{(Y)}\) varies with the compliance fraction even if compliance and treatment effects are uncorrelated, as has long been understood.

The key parameters in our simulation model are the notional tolerance values $A_1$ and $A_2$, which one can vary to produce any desired compliance configuration. It is plausible
that in a series of trials of a common experimental treatment, the parameters $A_1$ and $A_2$ would vary in the sense that as certainty about the treatment effect grows, subjects will be more likely to adhere to randomization. Compliance might also increase if the treatment's developers can, over time, ameliorate any negative side effects.

Ultimately, the selection of a trial estimand involves balancing considerations of scientific relevance and potential sensitivity to assumptions. One might argue that it is more important to know CACE, because it refers to the treatment effect among compliers, eliminating from consideration those who would not take the treatment if offered. Yet the latent subset of compliers itself may change depending on many factors. For example, an unproven treatment that causes unpleasant side effects may elicit poor compliance in an early trial. Once the treatment's positive effects become well established, or researchers are able to mitigate its side effects, compliance may improve substantially, rendering earlier estimates of CACE suspect.

The validity of instrumental variables estimation depends on the five assumptions listed above, and although some are plausible in many trials (e.g., SUTVA and randomization), others are often suspect (the exclusion-restriction). In any event, our simulations and real data analysis show clearly that CACE can also vary from trial to trial, depending on the level of compliance and its association with outcomes. The only evidence for this will come from observing such variability. Clearly, the anticipated level of compliance in a trial should inform the selection of the primary trial estimand.

An alternative view asserts that the most compelling estimand is PACE, or the average treatment effect could every patient be induced to comply. One cannot estimate this parameter directly in trials with noncompliance, but it is possible to create nonparametric bounds for it using linear programming techniques [3]. The width of these bounds depends on the proportion of noncompliance [57], and in practice the intervals are commonly too wide to provide useful information. Only studies with near-perfect compliance can avoid these redundancies.
A.1. Estimation of the unknown parameters with a binary outcome

Assume a population of \( N \) units, where for each unit \( i \) there is an associated vector \((Y_i, G_i) \in \{0, 1\} \times \{0, 1\}\) for the outcome of interest and its corresponding missingness indicator. The parameter of interest is \( \Pr[Y = 1] \). In Section 2.2.1 we assume the two link functions to be logit functions. Suppose observing a random sample of \( Y \) values of size \( n \) with \( n_m \) missing and \( n_o \) observed, where \( n = n_m + n_o \). Then, the maximum likelihood estimates (MLEs) for the probability of missing and \( \Pr[Y = 0|G = 1] \) can be estimated by the proportion of missing in the sample,

\[
\hat{p}_m = \frac{n_m}{n}
\]

and

\[
\hat{\mu}_c = 1 - \frac{1}{n_o} \sum_{i:g_i=1} y_i.
\]

Following the first two steps in Section 2.2.2 with fixed \( \pi_0, \gamma_1, \beta_1 \),

\[
\Pr[G = 0] = \frac{\pi_0}{1 + \exp(\gamma_0)} + \frac{1 - \pi_0}{1 + \exp(\gamma_0 + \gamma_1)}.
\]

(A.1)

Plug \( \hat{p}_m \) into Equation (A.1) with fixed \( \pi_0, \gamma_1 \) to obtain

\[
\exp(\gamma_0) = \frac{-[(\hat{p}_m - \pi_0) \exp(\gamma_1) + \hat{p}_m + \pi_0 - 1] + C_0}{2 \exp(\gamma_1) \hat{p}_m}.
\]
where $C_0 = \sqrt{[(\hat{p}_m - \pi_0) \exp(\gamma_1) + \hat{p}_m + \pi_0 - 1]^2 + 4 \exp(\gamma_1)\hat{p}_m(1 - \hat{p}_m)}$. Then

$$
\Pr[Y = 0|G = 1] = \frac{w}{1 + \exp(\beta_0)} + \frac{1 - w}{1 + \exp(\beta_0 + \beta_1)}, \quad (A.2)
$$

where

$$
w = \Pr[U = 0|G = 1] = \frac{\Pr[G = 1|U = 0]\Pr[U = 0]}{\Pr[G = 1|U = 0]\Pr[U = 0] + \Pr[G = 1|U = 1]\Pr[U = 1]}
= \pi_0 \left[ \pi_0 + \frac{\exp(\gamma_1)[1 + \exp(\gamma_0)]}{1 + \exp(\gamma_0 + \gamma_1)}(1 - \pi_0) \right]^{-1},
$$

and substituting $\pi_0, \gamma_1$ and $\hat{\gamma}_0$ to estimate $w$, denoted as $\hat{w}$. Plugging $\hat{w}$ and $\hat{\mu}_c$ into Equation (A.2) gives

$$
\exp(\hat{\beta}_0) = -\frac{[(\hat{\mu}_c - \hat{w}) \exp(\beta_1) + \hat{\mu}_c + \hat{w} - 1] + C_1}{2 \exp(\beta_1)\hat{\mu}_c},
$$

where $C_1 = \sqrt{[(\hat{\mu}_c - \hat{w}) \exp(\beta_1) + \hat{\mu}_c + \hat{w} - 1]^2 + 4 \exp(\beta_1)\hat{\mu}_c(1 - \hat{\mu}_c)}$. 

58
A.2. Proof of Equation 2.8

\[
E[Y|G = 0] - E[Y|G = 1] = E[Y|U = 1, G = 0]\Pr[U = 1|G = 0] + \\
E[Y|U = 0, G = 0]\Pr[U = 0|G = 0] - \\
E[Y|U = 1, G = 1]\Pr[U = 1|G = 1] - \\
E[Y|U = 0, G = 1]\Pr[U = 0|G = 1] \\
= E[Y|U = 1]\Pr[U = 1|G = 0] + \\
E[Y|U = 0](1 - \Pr[U = 1|G = 0]) - \\
E[Y|U = 1]\Pr[U = 1|G = 1] - \\
E[Y|U = 0](1 - \Pr[U = 0|G = 1]) \\
= E[Y|U = 1](\Pr[U = 1|G = 0] - \Pr[U = 1|G = 1]) - \\
E[Y|U = 0](\Pr[U = 1|G = 0] - \Pr[U = 1|G = 1]) \\
= (E[Y|U = 1] - E[Y|U = 0]) \\
\times (\Pr[U = 1|G = 0] - \Pr[U = 1|G = 1])
\]

A.3. A categorical confounder

A.3.1. Bounding inequality with categorical confounder

For the difference scale, MD_{YU} and MD_{UG} have been defined in the same notation from the above section. Without loss of generality, define level m of U to minimize E[Y|U = u_j]. To simplify the derivation, define |E[Y|G = 0] - E[Y|G = 1]| as D. It has been derived that |E[Y] - E[Y|G = 1]| = D \times \Pr[G = 0] and the only formula required to be derived in
terms of the two defined relations is $D$.

$$D = \sum_{i=1}^{m} E[Y|U = u_i]Pr[U = u_i|G = 0] - \sum_{i=1}^{m} E[Y|U = u_i]Pr[U = u_i|G = 1]$$

$$= \sum_{i=1}^{m-1} E[Y|U = u_i]Pr[U = u_i|G = 0] + E[Y|U = u_m][1 - \sum_{i=1}^{m-1} Pr[U = u_i|G = 0]] - \{ \sum_{i=1}^{m-1} E[Y|U = u_i]Pr[U = u_i|G = 1] + E[Y|U = u_m][1 - \sum_{i=1}^{m-1} Pr[U = u_i|G = 1]] \}$$

$$= \sum_{i=1}^{m-1} [E[Y|U = u_i] - E[Y|U = u_m]]Pr[U = u_i|G = 0] - \sum_{i=1}^{m-1} [E[Y|U = u_i] - E[Y|U = u_m]]Pr[U = u_i|G = 1]$$

$$\leq MD_{YU} \sum_{i=1}^{m-1} [Pr[G = u_i|G = 0] - Pr[G = u_i|G = 1]]$$

$$\leq [(m - 1)MD_{YU}MD_{UG}]$$

This is Inequality (2.16).

For the ratio scale, $ER_{YU(i)}$, $RR_{UG(i)}$, $MR_{YU}$ and $MR_{UG}$ have been defined in the same way with all of them assumed to be greater than 1. Without loss of generality, let

$$E[Y|U = u_1] = \min_i E[Y|U = u_i], \ E[Y|U = u_m] = \max_i E[Y|U = u_i].$$
Denote $R$ as $\frac{\mathbb{E}[Y|G=0]}{\mathbb{E}[Y|G=1]}$. Then

$$R = \frac{\sum_{i=1}^{m} \mathbb{E}[Y|U = u_i] \Pr[G = u_i|G = 0]}{\sum_{i=1}^{m} \mathbb{E}[Y|U = u_i] \Pr[G = u_i|G = 1]}$$

$$= \frac{\text{MR}_{YU} \Pr[U = u_1|G = 0] + \sum_{i=2}^{m-1} \text{ER}_{YU(i)} \Pr[G = u_i|G = 0] + \Pr[U = u_m|G = 0]}{\text{MR}_{YU} \Pr[U = u_1|G = 1] + \sum_{i=2}^{m-1} \text{ER}_{YU(i)} \Pr[G = u_i|G = 1] + \Pr[U = u_m|G = 1]}$$

$$= \frac{\sum_{i=1}^{m-1} (\text{ER}_{YU(i)} - 1) \Pr[G = u_i|G = 0] + 1}{\sum_{i=1}^{m-1} (\text{ER}_{YU(i)} - 1) \Pr[G = u_i|G = 1] + 1}$$

$$\geq \frac{(\text{MR}_{YU} - 1) \Pr[U = u_1|G = 0] + \sum_{i=2}^{m-1} C_{i0} + 1}{(\text{MR}_{YU} - 1) \Pr[U = u_1|G = 0] + 1}$$

$$\geq \frac{\text{MR}_{YU} + \text{RR}_{UG(1)} - 1}{\text{MR}_{YU} \text{RR}_{UG(1)}}$$

$$\geq \frac{\text{MR}_{YU} + \text{MR}_{UG} - 1}{\text{MR}_{YU} \text{MR}_{UG}}.$$

where $C_{i0} = (\text{RR}_{YU(i)} - 1) \Pr[G = u_i|G = 0]$.

With all the relative ratios greater than 1, $R < 1$:

$$\frac{\text{MR}_{YU} + \text{MR}_{UG} - 1}{\text{MR}_{YU} \text{MR}_{UG}} \leq R \leq 1$$

We can restate the bounds as

$$|R - 1| \leq \frac{(\text{MR}_{YU} - 1)(\text{MR}_{UG} - 1)}{\text{MR}_{YU} \text{MR}_{UG}},$$

giving Inequality (2.18). When some of the relative ratios are greater than 1 and others are less than 1, the derivation is more complicated, but similar tricks of finding a bounding value for the ratio of $\mathbb{E}[Y]$ and $\mathbb{E}[Y|G = 1]$ could be considered.
A.3.2. MinNI for a categorical confounder

The MinNI on the risk difference scale from Inequality (2.17) for \((MD_{YU}, MD_{UG})\) is, for a continuous outcome,

\[
\left( \max \left\{ \frac{k\sigma_{Y|G=1}}{(m-1)Pr[G=0]}, \sqrt{\frac{k\sigma_{Y|G=1}}{(m-1)Pr[G=0]}}, \min \left\{ 1, \sqrt{\frac{k\sigma_{Y|G=1}}{(m-1)Pr[G=0]} \right\} \right), \right.
\]

but for a binary outcome,

\[
\left( \sqrt{\frac{k\sigma_{Y|G=1}}{(m-1)Pr[G=0]}}, \sqrt{\frac{k\sigma_{Y|G=1}}{(m-1)Pr[G=0]} \right),
\]

where \(k\sigma_{Y|G=1} \leq (m-1)Pr[G=0]\). The MinNI on the risk ratio scale from Inequality (2.19) for \((MR_{YU}, MR_{UG})\) is

\[
\left( \frac{1}{1 - \sqrt{\frac{kCV_{Y|G=1}}{Pr[G=0]}}, \frac{1}{1 - \sqrt{\frac{kCV_{Y|G=1}}{Pr[G=0]} \right),
\]

where \(0 \leq kCV_{Y|G=1} < Pr[G=0] \leq 1\), which is the same as the binary case.
APPENDIX B
APPENDIX of CHAPTER 3

B.1. Appendix for Formulas in Calculation of ISNI

The general score equations obtained from the first step are,

\[ \frac{\partial l}{\partial \theta} \bigg|_{\nu=0} = \sum_{i=1}^{N} \left\{ h_{i} g_{i} \frac{\partial}{\partial \theta} \left[ \ln f_{\theta}^{Y_{i} | X_{i}, Z_{i}} (y_{i} | x_{i}, z_{i}) \right] + (1 - h_{i}) g_{i} \frac{\partial}{\partial \theta} \left[ \ln f_{\theta, \beta}^{Y_{i} | Z_{i}} (y_{i} | z_{i}) \right] \right\} = 0 \]

\[ \frac{\partial l}{\partial \beta} \bigg|_{\nu=0} = \sum_{i=1}^{N} \left\{ h_{i} \frac{\partial}{\partial \beta} \left[ \ln f_{\beta}^{X_{i} | Z_{i}} (x_{i} | z_{i}) \right] + (1 - h_{i}) g_{i} \frac{\partial}{\partial \beta} \left[ \ln f_{\theta, \beta}^{Y_{i} | Z_{i}} (y_{i} | z_{i}) \right] \right\} = 0 \]

where

\[ f_{\theta, \beta}^{Y_{i} | Z_{i}} (y_{i} | z_{i}) = \int f_{\theta}^{Y_{i} | X_{i}, Z_{i}} (y_{i} | u, z_{i}) f_{\beta}^{X_{i} | Z_{i}} (u | z_{i}) du. \]

The second factor of ISNI is,

\[ \begin{bmatrix} \nabla^2 l_{\theta \theta} \\ \nabla^2 l_{\beta \theta} \end{bmatrix} \bigg|_{\nu=0} = \begin{bmatrix} \nabla^2 l_{\theta \xi_1} & \nabla^2 l_{\theta \xi_2} & \nabla^2 l_{\theta \gamma_1} & \nabla^2 l_{\theta \gamma_2} \\ \nabla^2 l_{\beta \xi_1} & \nabla^2 l_{\beta \xi_2} & \nabla^2 l_{\beta \gamma_1} & \nabla^2 l_{\beta \gamma_2} \end{bmatrix} \bigg|_{\nu=0}. \]
For simplification, assume the two missingness mechanisms are logistic functions.

The general formulas for these second derivatives in ISNI,

\[
\nabla^2 l_{\theta \theta} \big|_{\nu=0} = \sum_{i=1}^{N} \left\{ h_i g_i \frac{\partial^2}{\partial \theta^2} \left[ \ln f_{\theta, i}^{X_i, Z_i}(y_i|x_i, z_i) \right] + (1 - h_i) g_i \frac{\partial^2}{\partial \theta^2} \left[ \ln f_{\theta, \beta}^{Y_i|Z_i}(y_i|z_i) \right] \right\}
\]

\[
\nabla^2 l_{\theta \beta} \big|_{\nu=0} = \sum_{i=1}^{N} \left\{ (1 - h_i) g_i \frac{\partial^2}{\partial \theta \partial \beta} \left[ \ln f_{\theta, \beta}^{Y_i|Z_i}(y_i|z_i) \right] \right\}
\]

\[
\nabla^2 l_{\beta \beta} \big|_{\nu=0} = \sum_{i=1}^{N} \left\{ h_i \frac{\partial^2}{\partial \beta^2} \left[ \ln f_{\beta, \gamma}^{X_i|Z_i}(x_i|z_i) \right] + (1 - h_i) g_i \frac{\partial^2}{\partial \beta^2} \left[ \ln f_{\theta, \beta}^{Y_i|Z_i}(y_i|z_i) \right] \right\}
\]

\[
\nabla^2 l_{\theta \xi} \big|_{\nu=0} = \sum_{i=1}^{N} (1 - h_i) g_i (1 - q_i) \frac{\partial}{\partial \theta} E[X_i^{mis}|Y_i, Z_i]
\]

\[
\nabla^2 l_{\beta \xi} \big|_{\nu=0} = \sum_{i=1}^{N} \left\{ (1 - h_i) g_i (1 - q_i) \frac{\partial}{\partial \beta} E[X_i^{mis}|Y_i, Z_i] - (1 - h_i)(1 - g_i)q_i \frac{\partial}{\partial \beta} E[X_i^{mis}|Z_i] \right\}
\]

\[
\nabla^2 l_{\xi \xi} \big|_{\nu=0} = - \sum_{i=1}^{N} \left\{ h_i (1 - g_i) q_i \frac{\partial}{\partial \theta} E[Y_i^{mis}|X_i, Z_i] + (1 - h_i)(1 - g_i)q_i \frac{\partial}{\partial \theta} E[Y_i^{mis}|Z_i] \right\}
\]

\[
\nabla^2 l_{\theta \gamma} \big|_{\nu=0} = - \sum_{i=1}^{N} (1 - h_i) g_i r_i \frac{\partial}{\partial \theta} E[X_i^{mis}|Y_i, Z_i]
\]

\[
\nabla^2 l_{\beta \gamma} \big|_{\nu=0} = - \sum_{i=1}^{N} \left\{ (1 - h_i) g_i r_i \frac{\partial}{\partial \beta} E[X_i^{mis}|Y_i, Z_i] + (1 - h_i)(1 - g_i) r_i \frac{\partial}{\partial \beta} E[X_i^{mis}|Z_i] \right\}
\]

\[
\nabla^2 l_{\theta \gamma} \big|_{\nu=0} = \sum_{i=1}^{N} \left\{ h_i (1 - g_i)(1 - r_i) \frac{\partial}{\partial \theta} E[Y_i^{mis}|X_i, Z_i] - (1 - h_i)(1 - g_i) r_i \frac{\partial}{\partial \theta} E[Y_i^{mis}|Z_i] \right\}
\]

\[
\nabla^2 l_{\beta \gamma} \big|_{\nu=0} = - \sum_{i=1}^{N} (1 - h_i)(1 - g_i) r_i \frac{\partial}{\partial \beta} E[Y_i^{mis}|Z_i]
\]

where \( q_i = \frac{\exp \{ \xi_0 + \xi_3 z_i + \xi_4 h_i \} }{1 + \exp \{ \xi_0 + \xi_3 z_i + \xi_4 h_i \} } \), \( r_i = \frac{\exp \{ \gamma_0 + \gamma_3 z_i \} }{1 + \exp \{ \gamma_0 + \gamma_3 z_i \} } \) and

\[
E[X_i^{mis}|Y_i, Z_i] = \int u f_{\theta}^{Y_i|X_i, Z_i}(y_i|u, z_i) f_{\beta}^{X_i|Z_i}(u|z_i) du\]

\[
\frac{f_{\theta, \beta}^{Y_i|Z_i}(y_i|z_i)}{f_{\theta, \beta}^{Y_i|Z_i}(y_i|z_i)}
\]


