Local Sensitivity to Nonignorable Missingness in the Overdispersed Count Outcome Using Negative Binomial Model

by

Bocheng Jing

M.Sc., Duke University, 2014
B.Sc., University of California, Davis, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Master of Science Program
Faculty of Health Sciences

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Bocheng Jing**
Degree:	**Master of Science**
Thesis title:	**Local Sensitivity to Nonignorable Missingness in the Overdispersed Count Outcome Using Negative Binomial Model**
Committee:	**Chair:** Lawrence McCandless Professor, Health Sciences

Hui Xie
Supervisor
Professor, Health Sciences

X. Joan Hu
Committee Member
Professor, Statistics and Actuarial Science

Yi Qian
Committee Member
Associate Professor, Business
University of British Columbia

Rachel Fouladi
Examiner
Associate Professor, Psychology

Abstract

Count outcomes with missing observations are a common occurrence in clinical, medical, and psychological research. Typically, such data is analyzed using statistical methods predicated on the missing at random assumption (MAR). However, verifying MAR from the data at hand can often be challenging, and an incorrect assumption could lead to biased results. The Index of Local Sensitivity to Nonignorability (ISNI) method provides a straightforward and user-friendly solution for sensitivity analysis. The corresponding R package, "isni", incorporates the Poisson model for count data. Yet, when the data exhibit overdispersion, using the Poisson model may not be suitable. To address this, we have developed an ISNI index for the Negative Binomial model, often employed for count outcomes in instances of overdispersion. We conducted simulation studies to explore the effects of the degree of overdispersion on the sensitivity to nonignorability, as well as the association between different levels of missing proportions and sensitivity to nonignorability. Our newly developed R function, isniglm.nb(), enables the implementation of the method for the negative binomial model. We demonstrate the application of this negative binomial ISNI method using a real-world dataset drawn from clinical research.

Keywords: Missing Data; ISNI Index; Negative Binomial Model; Count outcome; Simulation Study

Acknowledgements

I would like to express my deepest and most heartfelt gratitude to my supervisor, Professor Hui Xie, for his invaluable guidance, meticulous attention to detail, and unwavering support throughout the course of this research. His expertise, patience, and insightful feedback have been the pillars of strength and motivation behind the success of this thesis.

Additionally, I am profoundly thankful to Professors Yi Qian and Daniel F. Heitjan for their invaluable contributions to the ISNI tutorial paper. Their astute critiques and constructive suggestions have significantly enriched my understanding and approach towards the complex realm of missing data research.

I extend my special thanks to Professor X. Joan Hu for her astute suggestions and insightful comments which have greatly enhanced the quality of my project. Her expertise and thoughtful guidance have been instrumental in refining my research approach. I am also incredibly grateful to her team, comprising Charlie Zhou, Angela Chen, Trevor Thomson, Yi Xiong, Linwan Xu, and Ken Peng. Each team member's unique perspective and valuable input have enriched my research experience and contributed significantly to the development of my project.

My journey would not have been as rewarding and enjoyable without the camaraderie and support of my friends and colleagues, Kai Zhao, Yuetong Zhou, Lulu Guo, and Rashed Hoque. Their constant encouragement, intellectual discussions, and unwavering support in times of challenge have been a source of immense comfort and motivation. I am truly blessed to have such a supportive group of friends who have stood by me throughout this journey.

I am deeply indebted to the Simon Fraser Faculty of Health Science for their generous funding and for providing a stimulating intellectual environment, state-of-the-art facilities, and invaluable resources. The opportunity to work in such an inspiring academic setting has played a pivotal role in the successful completion of my research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The missing data problem

Missing data is a pervasive issue in clinical and psychological research. A systematic review of randomized controlled trials (RCTs), widely recognized as the gold standard for causal inference, revealed that 95% of the 77 reviewed articles published in 2013 contained varying levels of missing outcomes, with a median missing percentage of 9% (ranging from 0% to 70%) (Bell et al., 2014). Weiss et al. (2016) reported that 78% of 262 studies published in major epidemiological journals in 2010 disclosed missing data. However, the study noted an alarming deficiency in sensitivity analyses in these studies (Weiss et al., 2016). Moreover, systematic reviews highlight the problem of missing data in psychological studies. For instance, 48% of 1,666 studies published in 11 leading psychology and education journals from 1998 to 2004 reported missing data (Peng et al., 2006). Similarly, in the Journal of Counseling Psychology in 2008, 80% of the 46 quantitative research articles reported missing data (Schlomer et al., 2010). Jeličić et al. (2009) found that 57 out of 100 reviewed studies in developmental psychology either acknowledged missing data or showed clear evidence of its presence.

The issue of missing data poses significant threats to study validity in two key ways: Firstly, missing data can lead to a reduced sample size, consequently undermining the study's power and efficiency. Secondly, if there is a correlation between the data values and the likelihood of them being missing, the observed data may not be representative of the intended population, leading to biased data analysis.

To understand this bias, it is essential to recognize that the missing data itself is an integral part of the data and should ideally be treated as such. Rubin (1976) proposed a condition, termed *missing at random (MAR)*, which is critical for determining when conventional analyses of incomplete data sets yield valid results. Roughly, the missing data is considered MAR when the missing observations, if known, would provide no predictive value for their own missingness.

In particular, MAR ensures the validity of two statistical methods commonly used in clinical research: Maximum likelihood estimation and multiple imputation (Schafer and Graham, 2002).

These methods allow analysts to incorporate all observed data elements in a way that mitigates bias and guarantees efficiency. They enable analyses of incomplete data sets in various applications, including regression analysis (Schafer and Graham, 2002; Qian and Xie, 2011; Chen et al., 2011; Enders et al., 2020; Lüdtke et al., 2020), matched sampling and propensity score analysis (Qian, 2007; Cham and West, 2016), moderation analysis (Q. Zhang and Wang, 2017), structural equation modeling (T. Lee and Shi, 2021), data fusion (Qian and Xie, 2014), data privacy and disclosure control (Qian and Xie, 2015), among others.

1.2 Ignorability reduces the complexity of analyses

To derive valid inferences from incomplete data, one often needs to estimate not only the model for the outcomes but also a secondary model that explicates how the data became missing. This is known as the *missing data mechanism (MDM)* (Little and Rubin, 2019). However, as MDMs are often scientifically uninteresting and difficult to estimate, we aim to find methods to circumvent them. In other words, we hope that the MDM is ignorable so that we can draw valid inferences about the relevant model parameters without the need to estimate the MDM. This approach essentially acts as if the intention was always to collect only the observed data and not the data that is missing.

When basing inferences on the likelihood or Bayesian estimation, the MDM is ignorable if the data are Missing At Random (MAR) and the parameters guiding the generation of the underlying complete data are distinct from those governing the MDM. Hence, the essence of MAR is its capacity to assure *ignorability* (Heitjan and Rubin, 1991). MAR applies when, given the observed data values, missingness indicators and unobserved values are independent (Rubin, 1976; Little and Rubin, 2019). For univariate analysis, MAR is satisfied if there is no association between the outcome variable and the probability of an observation being missing. In regression analysis with potentially missing outcomes, if missingness depends solely on predictors (or "independent variables"), the data are considered MAR. In a longitudinal study, if the chance of a subject dropping out at time t depends only on baseline predictors and preceding values of the outcome, the data are again deemed MAR.

Programs like `SAS Proc Mixed`, which claim to accommodate missing data, are underpinned by the MAR assumption. However, this assumption is not inconsequential; treating the MDM as ignorable when it is not can lead to severely biased results (Blankers et al., 2010; Shin et al., 2017; T. Lee and Shi, 2021). Consequently, several authors advocate assessing the robustness of standard analyses to alternative missing data assumptions as a crucial aspect of study reporting (Little, D'Agostino, et al., 2012).

1.3 Models for missing not at random (MNAR)

In cases with Missing Not At Random (MNAR) data, a joint model for the outcome Y_i and missing indicator G_i is often employed. Specifically, $G_i = 1$ if Y_i is missing and $G_i = 0$ otherwise. Three main approaches under MNAR account for nonignorable missingness: pattern-mixture, shared-parameter, and selection models (Molenberghs et al., 2014).

In pattern-mixture models, the joint distribution is decomposed into the conditional distribution of Y_i given G_i and the distribution of the missing indicator G_i conditioned on the covariates X (Little, 1993; Molenberghs et al., 2014). The overall distribution combines both observed and unobserved cases (Galimard et al., 2018; Leurent et al., 2018). Notably, $Y_i|G_i$ is unidentifiable due to some outcomes being unobserved. Restrictions are imposed on these unidentified parameters, linking them either to other missing data patterns or to parameters identified from observed data (Little, 1993). Sensitivity analysis, which involves varying these restrictions, is key to pattern-mixture models (Curran et al., 2004). One advantage of this approach is that the missing data mechanism can be easily derived using Bayes' rule once these restrictions are set (Molenberghs et al., 2014). However, this comes at the cost of increased complexity in parameter estimation, as it requires marginalizing the outcome distribution over the missing patterns.

Shared-parameter models introduce random effects, denoted by b, into the decomposition of the joint distribution into $Y_i|b_i$ and $G_i|b_i, Y_i$. These random effects create a connection between the outcome Y_i and the missing data indicator G_i. Each subject has a unique random effect influencing both the outcome and the likelihood of data being missing (Molenberghs et al., 2014). With these random effects in place, the outcome and missing data mechanisms are assumed to be independent (Creemers et al., 2010; Rizopoulos et al., 2008). The framework for sensitivity analysis in shared-parameter models was first proposed by Rizopoulos et al. (2008), and later extended by Creemers et al. (2010) to specifically address longitudinal data.

In the selection model, the joint distribution is partitioned into the marginal distribution of Y_i and the conditional distribution of $G_i|Y_i$. The marginal distribution focuses on the observed data, while the probability of missing data is conditional on unobserved outcomes (Molenberghs et al., 2014). Initially introduced by Heckman (1979), the selection model models the missing probability $P(G_i)$ based on the unobserved outcome. While the selection model offers the advantage of directly utilizing the marginal distribution of observed outcomes, it also presents several challenges. First, determining restrictions for the conditional model is ambiguous, making it problematic to identify parameters when these restrictions change (Glonek, 1999). Second, finding a closed form for the conditional distribution of missing data is often elusive, complicating the model fitting process (Molenberghs et al., 2014). Lastly, conducting a sensitivity analysis can be challenging, particularly when parameters for X are inferred from observed data. Recent advancements have employed semi-parametric approaches for sensitivity analysis within the selection model (Tsiatis, 2006). The local sensitivity method further expands this sensitivity analysis approach.

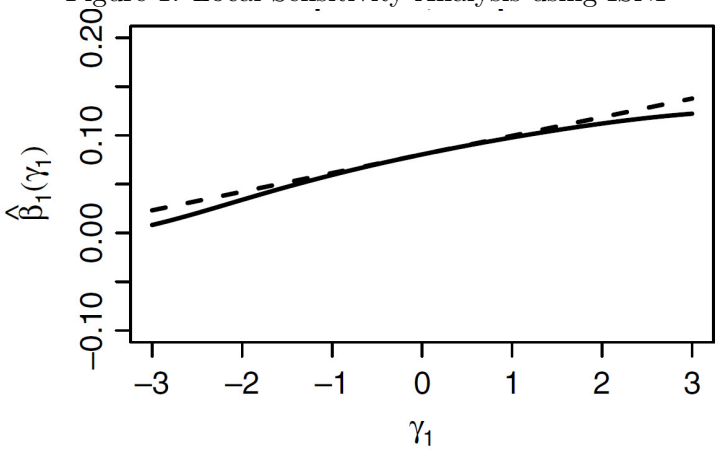
1.4 A local sensitivity analysis to nonignorability

A robust test for MAR (Missing at Random) cannot be reliably conducted using the available data, as we do not have access to the values of the missing items (Kenward, 1998; Little and Rubin, 2019). However, by relaxing the MAR assumption, we can observe its effects on the estimates of regression coefficients. The ISNI (Index of Sensitivity to Nonignorability) method, a straightforward approach proposed by Troxel et al. (2004) (akin to a method earlier suggested by Copas and Li (1997)), allows for sensitivity analysis of this kind.

The essential concept of the ISNI analysis is outlined as follows: Initially, one specifies a model for the outcome in the absence of missing data. Then, an MDM (missing data mechanism) is specified in which the probability of missingness is dependent on the outcome subject to missingness. This MDM should be indexed by a *nonignorability parameter*, which, when set to 0, gives rise to a MAR model. For instance, the MDM could be a logistic regression of the missingness probability on the outcome measure. The slope coefficient in this logistic model becomes a nonignorability parameter, as the missingness is independent of the outcome when it equals 0.

Consider the estimation of the parameter of interest in the outcome model, keeping the non-ignorability parameter fixed at some non-zero value. By calculating this estimate at a range of non-zero values (non-MAR MDMs), a curve mapping the estimated outcome parameter against the nonignorability parameter is created. This is the essence of the sensitivity analysis. If the curve is flat near the MAR model, the estimation is considered robust to local departures from ignorability. Conversely, if it is steep, the estimation is deemed sensitive to nonignorability. Given its crucial role as a measure of sensitivity, the slope of the curve at the MAR model is termed the *Index of Local Sensitivity to Nonignorability (ISNI)* (Figure 1).

Figure 1: Local Sensitivity Analysis using ISNI



Note: The solid curve plots the outcome model parameter estimate $\widehat{\beta}_1(\gamma_1)$ against the nonignorability parameter γ_1. ISNI is the slope of the tangent line (dashed line) for the curve at $\gamma_1 = 0$ (the MAR model). Reproduced from Figure 2 in Xie and Qian (2012).

Conveniently, the ISNI method hinges on terms that can be easily computed during the estimation of the MAR model. As such, an approximate sensitivity analysis can be performed without the need to estimate a series of nonignorable models. The ISNI-based sensitivity analysis offers several beneficial features, as described below.

First, ISNI analysis produces parsimonious sensitivity index measures and is simple to conduct. ISNI adopts a local sensitivity analysis approach (Copas and Li, 1997; Verbeke et al., 2001) that measures the rate of change in parameter estimates when the MDM departs from MAR. ISNI computation avoids estimating any nonignorable models; only the readily available MAR estimates are required to compute ISNI. Second, ISNI is amenable to *systematic* sensitivity analysis with minimal data requirements. ISNI is applicable to a wide variety of statistical models frequently used in psychology research, including linear regression, ANOVA, ANCOVA, binary regression, Poisson regression, other generalized linear models (Troxel et al., 2004), marginal multivariate models (e.g., repeated measures ANOVA) (Ma et al., 2005), multi-level generalized linear mixed effect models (also known as hierarchical models or latent growth curve models) for longitudinal/clustered data (Xie, 2008; J. K. Murphy et al., 2020; Catherine et al., 2020; Woodward et al., 2021), time-to-event models (J. Zhang and Heitjan, 2006; T. Liu and Heitjan, 2012), and Bayesian inference (J. Zhang and Heitjan, 2007; Xie, 2009). Moreover, ISNI requires neither monotone missing data patterns nor intermittent missingness to be ignorable and can handle nonignorable non-monotone missing data and complex reasons in both experimental (Xie, 2012; Xie and Heitjan, 2004) and observational studies (Xie and Qian, 2012; Xie, Qian, and Qu, 2011) as well as high-dimensional missingness problems occurring in new types of psychological data. Thus, ISNI permits a systematic screening for sensitivity without restricting the sizes, numbers, or types of analyses performable (Gao et al., 2016; Yuan et al., 2020; Schneider et al., 2021; Razzaq et al., 2022). Third, the method is made *accessible* via the `R` package `isni`. The lack of user-friendly software can be a crucial barrier to the adoption of the ISNI method. The recent implementation of ISNI in the `R` package `isni` aims to alleviate this obstacle and facilitate the adoption of the method by quantitative researchers and data scientists in psychology and other related fields (Xie, Gao, et al., 2018). The program's default setting computes ISNI with minimal coding difference from MAR analysis, requiring neither the creation of new variables nor the specification of new models beyond those required for regular MAR analysis(Jing et al., 2023).

1.5 Missing data in count outcome

Count outcomes are commonly encountered in clinical, medical, and psychological research. They represent discrete and non-negative numerical values that indicate the frequency or count of an event occurring within a specified time or space. For instance, comparing the number of high-risk medications taken by dementia patients living alone versus those living with others can provide insights into medication management. Exploring the occurrence of infections in systemic lupus

erythematosus(SLE) patients and its association with other risk factors is of significant research interest for developing preventative interventions in the field of arthritis. Additionally, comparing the number of activities of daily living (ADL) difficulties experienced by individuals after major cardiac surgeries can inform decision-making in an aging population.

Count outcomes are frequently collected via nationally representative surveys or self-reported measurements. In the case of ADL, activities were collected via the Health and Retirement Study (HRS). Nevertheless, such count outcomes from survey studies often come with nonresponses, leading to missing outcome data. In various instances, researchers assumed missing at random and carried out the subsequent analysis (Brown, Diaz-Ramirez, Boscardin, S. J. Lee, and Steinman, 2017, Stineman et al., 2016, Lawrence et al., 2004, Brown, Diaz-Ramirez, Boscardin, S. J. Lee, Williams, et al., 2019). Moreover, the missing count outcomes can occur in situations where patients switch healthcare plans that do not provide available medication data. Similarly, if patients transfer out of the network, infections may no longer be reported through the provincial health portal.

When modeling count data outcomes, two commonly used regression methods are negative binomial regression and Poisson regression. These models are specifically designed to handle count data and account for the inherent characteristics and distributional assumptions of such data.

Poisson regression assumes that the mean and variance of the count data are equal. It is appropriate when the count data follow a Poisson distribution and exhibit no or minimal overdispersion. Negative binomial (NB) regression, on the other hand, is typically used when the count data exhibit overdispersion, in which the variance is greater than the mean. It allows for the modeling of count outcomes with excessive variability that cannot be adequately captured by the Poisson regression model.

Although the ISNI index has been developed for Poisson regression (Troxel et al., 2004), its adaptation for negative binomial (NB) regression, which deals with overdispersed count outcomes, is yet to be developed. The inclusion of the dispersion variable adds another level of complexity in estimating the covariance matrix for the negative binomial (NB) model, which is an integral component of the ISNI index. Furthermore, the extent to which the degree of dispersion affects the sensitivity of missing data has yet to be determined.

1.6 Objective and Outline

The objective of this thesis will be to develop the ISNI index specifically for the negative binomial (NB) regression. The thesis will be organized as follows: Chapter 2 will focus on specifying the nonignorable missing model in a general form and deriving the ISNI index based on this general form. Chapter 3 will delve into the specification of the NB regression model, joint likelihood estimation, and the derivation of the ISNI index for the NB regression. A comparison of sensitivity analysis between NB and Poisson regressions will be conducted, and simulation studies will be carried out to explore the impact of dispersion degree and various missing proportions on sensitivity.

Chapter 4 will present an application where the NB ISNI index will be utilized to examine the sensitivity of missing activities of daily living (ADL) measures in post-major surgery patients.

Chapter 2

ISNI Derivation in General Form

2.1 Specify the general outcome model and the selection model

Denote Y as the outcome variable, and X as the set of predictor variables. The parameter θ represents the conditional distribution of Y, given X, encapsulated by the density function $f_\theta(y_i|x_i)$ for independent subjects i = 1, ..., n. Without loss of generality, we assume $f_\theta(y_i|x_i)$ is a probability density function for a continuous outcome and a probability mass function for a discrete outcome. The missingness indicator, represented as G_i, is defined such that $G_i = 1$ (0) if y_i is missing (observed). In the context of nonignorable missingness, it is assumed that the probability of an observation being missing depends on both Y_i and a set of predictors s_i. It is important to note that s_i may overlap with x_i. We permit the probability of Y_i being missing to depend on the actual value of Y_i through a nonignorable parameter γ_1 via a logistic regression model, such as

$$P(G_i = 1|Y_i = y_i, S_i = s_i) = h(\gamma_0 s_i + \gamma_1 y_i)$$

For further notation, we also specify the equation of $P(G_i = 0|Y_i = y_i, S_i = s_i)$ and the derivative of the logistic regression,

$$P(G_i = 0|Y_i = y_i, S_i = s_i) = 1 - P(G_i = 1|Y_i = y_i, S_i = s_i) = 1 - h(\gamma_0 s_i + \gamma_1 y_i)$$

The joint density function, $f(Y, G)$, is given by

$$\begin{aligned} f(Y, G) &= f(Y) \times f(G|Y) \\ &= f_\theta(y^{\text{obs}}, y^{\text{mis}}) \times f_{\gamma_0,\gamma_1}(g|y^{\text{obs}}, y^{\text{mis}}) \end{aligned}$$

In this thesis, we examine $f_\theta(y^{\text{obs}}, y^{\text{mis}})$ for cross-sectional data. Therefore, the joint likelihood under this model, $L(\theta, \gamma_0, \gamma_1; y^{\text{obs}}, y^{\text{mis}}, g, x, s)$, can be specified as

$$L(\theta, \gamma_0, \gamma_1; y^{\text{obs}}, y^{\text{mis}}, g, x, s) = \prod_{i=1}^{n} \Big[f_\theta(y_i^{\text{obs}}|x_i)(1 - h(\gamma_0 s_i + \gamma_1 y_i^{\text{obs}})) \Big]^{1-g_i} \Big[\int_{\Omega_{y^{\text{mis}}}} f_\theta(y_i^{\text{mis}}|x_i) h(\gamma_0 s_i + \gamma_1 y_i^{\text{mis}}) dy_i^{\text{mis}} \Big]^{g_i}$$

The log-likelihood can be specified as

$$\begin{aligned} \ln L = \sum_{i=1}^{n} \Bigg\{ & (1 - g_i) \Big[\ln f_\theta(y_i^{\text{obs}}|x_i) + \ln(1 - h(\gamma_0 s_i + \gamma_1 y_i^{\text{obs}})) \Big] \\ & + g_i \ln \int_{\Omega_{y^{\text{mis}}}} f_\theta(y_i^{\text{mis}}|x_i) h(\gamma_0 s_i + \gamma_1 y_i^{\text{mis}}) dy_i^{\text{mis}} \Bigg\}, \end{aligned} \tag{2.1}$$

y^{obs} and y^{mis} represent the observed and missing components of Y, respectively. $\Omega_{y^{\text{mis}}}$ is the sample space of Y^{mis}. The parameters γ_0 and γ_1 denote the effects on the probability of missingness that stem from fully observed data and potentially unobserved data, respectively. In the logistic regression model, which is often utilized to model missing data indicators G, γ_0 corresponds to the set of coefficients for the fully observed missingness predictors. γ_1 corresponds to the coefficients for the outcomes that could potentially be unobserved. In essence, γ_1 is the *nonignorability parameter*. When the condition $\gamma_1 = 0$ is met, the Missing Data Mechanism (MDM) simplifies to the Missing At Random (MAR) scenario.

2.2 ISNI Definition and Derivation

2.2.1 ISNI Definition

The nonignorability parameter, denoted as γ_1, is generally challenging to estimate. Consequently, a method for sensitivity analysis involves evaluating $\widehat{\theta}(\gamma_1)$ — that is, the estimate of θ if we were aware of γ_1's value — across a reasonable range of γ_1 values. Due to the difficulty in computing $\widehat{\theta}(\gamma_1)$ for any non-zero γ_1, Troxel et al. (2004) suggested a Taylor Series linear approximation as follows:

$$\hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \left. \frac{\partial \widehat{\theta}(\gamma_1)}{\partial \gamma_1} \right|_{\gamma_1 = 0} \gamma_1,$$

In this context, $\hat{\theta}(0)$ denotes the maximum likelihood estimate under the MAR model; the derivative term is identified as the Index of Local Sensitivity to Nonignorability or ISNI. In the following, we exemplify the derivation of the ISNI via implicit differentiation. For a fixed γ_1, we differentiate the

likelihood function with respect to(θ, γ_0), and the MLE satisfies

$$\frac{\partial L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial(\theta, \gamma_0)} = 0.$$

Note that $\widehat{\theta}$ and $\widehat{\gamma}_0$ are implicit functions of γ_1, $\widehat{\theta}(\gamma_1)$ and $\widehat{\gamma_0}(\gamma_1)$. For simplicity, in the following derivation, we keep the notation as $\widehat{\theta}$ and $\widehat{\gamma_0}$. When we differentiate both sides with respect to γ_1, we obtain

$$\frac{\partial^2 L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial(\theta, \gamma_0)\partial\gamma_1} + \frac{\partial^2 L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial(\theta, \gamma_0)\partial(\theta, \gamma_0)} \times \frac{\partial L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial\gamma_1} = 0$$

Thus for the derivative of γ_1, we have

$$\frac{\partial L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial\gamma_1} = -[\frac{\partial^2 L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial(\theta, \gamma_0)\partial(\theta, \gamma_0)}]^{-1} \times \frac{\partial^2 L(\widehat{\theta}, \widehat{\gamma}_0, \gamma_1)}{\partial(\theta, \gamma_0)\partial\gamma_1}$$

ISNI is defined as the first derivative evaluated at $\gamma_1 = 0$,

$$\begin{aligned}
\text{ISNI} &= \frac{\partial(\widehat{\theta}(\gamma_1), \widehat{\gamma}_0(\gamma_1))}{\partial\gamma_1}\Big|_{\gamma_1=0} \\
&= -\begin{bmatrix} \nabla^2 L_{\theta,\theta} & \nabla^2 L_{\theta,\gamma_0} \\ \nabla^2 L_{\gamma_0,\theta} & \nabla^2 L_{\gamma_0,\gamma_0} \end{bmatrix}^{-1} \begin{bmatrix} \nabla^2 L_{\theta,\gamma_1} \\ \nabla^2 L_{\gamma_0,\gamma_1} \end{bmatrix},
\end{aligned}$$

where $\nabla^2 L$ denotes the partial derivative for two parameters $(p, q) \in (\theta, \gamma_0, \gamma_1)$, such that

$$\nabla^2 L_{p,q} = \frac{\partial^2 L(\theta(\gamma_1), \gamma_0(\gamma_1), \gamma_1)}{\partial p^T \partial q}\Big|_{\widehat{\theta}(0), \widehat{\gamma}_0(0), 0}.$$

Because the off-diagonal term, $\nabla^2 L_{\theta,\gamma_0}$, is 0, ISNI for $\widehat{\theta}$ simplifies as

$$\frac{\partial\widehat{\theta}(\gamma_1)}{\partial\gamma_1} = -\nabla^2 L_{\theta,\theta}^{-1} \nabla^2 L_{\theta,\gamma_1}.$$

2.2.2 Derive $\nabla^2 L_{\theta,\theta}$ and $\nabla^2 L_{\theta,\gamma_1}$

The comprehensive derivation of $\nabla^2 L_{\theta,\theta}$ and $\nabla^2 L_{\theta,\gamma_1}$ from the joint log-likelihood equation 2.1 unfolds as follows:

$$\begin{aligned}
\nabla_{\theta,\theta} &= -\frac{\partial^2}{\partial\theta^2}\ln Ł \\
&= -\frac{\partial}{\partial\theta}\sum_{i=1}^{n}(1-g_i)\frac{\partial \ln f_\theta(y_i|x_i)}{\partial\theta} + gi\frac{\frac{\partial}{\partial\theta}\int_\Omega f_\theta(y_i|x_i)h(\gamma_0 s_i+\gamma_1 y_i)dy_{mis}}{\int_\Omega f_\theta(y_i|x_i)h(\gamma_0 s_i+\gamma_1 y_i)dy_{mis}} \\
&= -\sum_{i=1}^{n}(1-g_i)\frac{\partial^2 \ln f_\theta(y_i|x_i)}{\partial\theta^2}
\end{aligned}$$

Note that -$\nabla_{\theta,\theta}$ is the Fisher Information from the complete cases hence -$\nabla^{-1}_{\theta,\theta}$ is the covariance matrix from the observed data under the MAR model.

$$\begin{aligned}
\nabla^2_{\theta,\gamma_1} &= \frac{\partial^2}{\partial\theta\partial\gamma_1}\ln Ł \\
&= \frac{\partial}{\partial\theta}\frac{\sum_{i=1}^{n}\left\{(1-g_i)\left[\ln f_\theta(y_i|x_i)+\ln[1-h(\gamma_0 s_i+\gamma_1 y_i)]\right]+g_i\ln\int_\Omega f_\theta(y_i|x_i)h(\gamma_0 s_i+\gamma_1 y_i)dy_{\text{mis}}\right\}}{\partial\gamma_1} \\
&= \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left\{(1-g_i)\frac{-h'(\gamma_0 s_i+\gamma_1 y_i)}{1-h(\gamma_0 s_i+\gamma_1 y_i)}+\frac{g_i\int_\Omega f_\theta(y_i|x_i)\frac{\partial}{\partial\gamma_1}h(\gamma_0 s_i+\gamma_1 y_i)dy_{\text{mis}}}{\int_\Omega f_\theta(y_i|x_i)h(\gamma_0 s_i+\gamma_1 y_i)dy_{\text{mis}}}\right\} \\
&= \sum_{i=1}^{n}\left\{\frac{\partial}{\theta}\frac{g_i\int_\Omega f_\theta(y_i|x_i)h'(\gamma_0 s_i+\gamma_1 y_i)dy_{\text{mis}}}{h(\gamma_0 s_i+\gamma_1 y_i)\int_\Omega f_\theta(y_i|x_i)dy_{\text{mis}}}\right\} \\
&= \sum_{i=1}^{n}\left\{g_i\frac{\frac{\exp(\gamma_0 s_i+\gamma_1 y_i)}{[1+exp(\gamma_0 s_i+\gamma_1 y_i)]^2}\frac{\partial}{\partial\theta}\int_\Omega y_i f_\theta(y_i|x_i)dy_{\text{mis}}}{\frac{\exp(\gamma_0 s_i+\gamma_1 y_i)}{1+\exp(\gamma_0 s_i+\gamma_1 y_i)}}\right\} \\
&= \sum_{i=1}^{n}\left\{g_i\left[1-h(\gamma_0 s_i+\gamma_1 y_i)\right]\frac{\partial E_\theta(y_i|x_i)}{\partial\theta}\right\} \\
&= \sum_{i=1}^{n}\left\{g_i\left[\frac{1}{1+\exp(\gamma_0 s_i+\gamma_1 y_i)}\right]\frac{\partial E_\theta(y_i|x_i)}{\partial\theta}\right\}
\end{aligned}$$

Here, $E_\theta(y_i|x_i)$ denotes the conditional mean of the outcome that is missing. $\nabla^2_{\theta,\gamma_1}$ signifies the degree of nonorthogonality between θ and γ_1. It's noteworthy that the calculation of this term merely requires the MAR estimates $\widehat{\theta}(0)$ and $\widehat{\gamma}_0(0)$, thus it can be readily computed using other data information.

Together, we derived ISNI as

$$\texttt{ISNI} = -\nabla_{\theta,\theta}^{-1}\nabla_{\theta,\gamma_1} = -\left[\sum_{i=1}^{n}(1-g_i)\frac{\partial^2 \ln f_\theta(y_i|x_i)}{\partial\theta^2}\right]^{-1}\sum_{i=1}^{n}\left\{g_i\left[\frac{1}{1+\exp(\gamma_0 s_i+\gamma_1 y_i)}\right]\frac{\partial E_\theta(y_i|x_i)}{\partial\theta}\right\}$$

Troxel et al. (2004) derived ISNI formulas for generalized linear regression models, now implemented in function `isniglm()` in the `R` package `isni` (Xie, Gao, et al., 2018). Thus one can now readily compute ISNI for linear regression, ANOVA, ANCOVA, binary logistic and probit regression, Poisson regression, and Gamma regression.

2.3 ISNI Interpretability

ISNI represents the rate of change in $\widehat{\theta}(\gamma_1)$ at $\gamma_1 = 0$ and therefore approximates the displacement in MLEs from the MAR estimate when γ_1 shifts from 0 to 1, $\hat{\theta}(1) \approx \hat{\theta}(0) + \texttt{ISNI}$. With the logit link for the MDM, the parameter $\gamma_1 = 1$ implies that a one-unit increment in y is associated with a 2.7-fold (exp(1)) increase in the odds of missingness. When outcomes have a natural scale, such as binary or categorical outcomes, the ISNI can be interpreted in this native scale. Specifically, if nonignorability is defined such that a 1-unit change in the outcome corresponds to a 2.7-fold increase in the odds of missingness, then the parameter estimate will shift by ISNI units.

For continuous y, the ISNI value is susceptible to changes in the unit of measurement, making its interpretation less straightforward. To enhance its interpretability, one can consider the impact of moderate nonignorability on the parameter estimates such that a one standard deviation increase in y would lead to a 2.7-fold increase in the odds of missingness. That is, one varies γ_1 from $-\frac{1}{\sigma_Y}$ to $\frac{1}{\sigma_Y}$, and then examines the range of estimates as

$$\widehat{\theta}(\gamma_1) \approx \widehat{\theta}(0) \pm \text{ISNI}(\widehat{\theta}) \times \gamma_1 = \widehat{\theta}(0) \pm \text{ISNI}(\widehat{\theta}) \times \frac{1}{\sigma_Y}, \tag{2.2}$$

This approach maintains the interpretability of ISNI while dealing with continuous outcomes.

In the original ISNI paper (Troxel et al., 2004), count outcomes were considered to be on a natural scale. However, unlike binary or categorical outcomes, which are bounded, count outcomes can theoretically be unbounded. Although count outcomes are measured on a natural scale (for instance, the count increases by 1 or the number of incidences increases by 1 unit within a certain period), the range of count outcomes can vary greatly. Consequently, a one-unit change within different outcome ranges may imply different degrees of probability of missingness. Therefore, we modify the definition from the original paper and scale the ISNI value by the standard deviation of the count outcome.

2.4 The Minimal Degree of Nonignorability (MinNI)

To determine the significance of a shift in the MLE estimate, one can compare the change in an MLE with its sampling error. In this context, a displacement of $\widehat{\theta}$ by one standard error (SE) is considered a significant degree of nonignorability,

$$\widehat{\theta}(\gamma_1) = \widehat{\theta}(0) \pm SE_\theta, \tag{2.3}$$

Contrasting equation 2.2 with equation 2.3, we observe that if $\frac{\text{ISNI}}{\sigma_Y}$ $/\text{SE}_\theta > 1$, it implies a high susceptibility of the model to nonignorability. We introduce the minimum degree of nonignorability (MinNI), also referred to as the c index in Troxel et al., 2004. MinNI is a scaling factor wherein a change of $\frac{1}{\text{MinNI}}$ standard deviations of Y corresponds with an odds ratio of 2.7 in the probability of being missing. Therefore, we can derive MinNI as:

$$\text{ISNI} \times \frac{1}{\frac{1}{\text{MinNI}}\sigma_Y}/\text{SE}_\theta = 1 \rightarrow \text{MinNI} = \left|\frac{\sigma_Y \text{SE}}{\text{ISNI}}\right|$$

The MinNI, as defined, is independent of the scale of the continuous or count outcome y. It represents the minimal nonignorability, expressed in 1/SD units, that would induce sensitivity. Therefore, a small MinNI implies that even a modest degree of nonignorability could lead to sensitivity, while a large MinNI suggests that only an extreme degree of nonignorability could result in sensitivity.

For instance, a MinNI of 10 signifies that, in order to change the MLE by one SE, the magnitude of nonignorability ($|\gamma_1|$) must be at least $\frac{1}{\frac{1}{10}\sigma_Y}$. This corresponds to a 0.1-SD unit change in Y leading to a 2.7-fold change in the odds for missingness. Given that such a minor variation in y is unlikely to alter the probability of missingness drastically, we view this degree of nonignorability as being too extreme to commonly occur in practice. As such, we consider inferences to be insensitive to nonignorability when MinNI is large.

Conversely, a MinNI of 0.1 indicates that, to modify the MLE by one SE, the degree of nonignorability ($|\gamma_1|$) must be no less than $\frac{1}{\frac{1}{0.1}\sigma_Y}$. This means that a 10-SD unit change in Y would equate to a 2.7-fold change in the odds for missingness. Given that a substantial change in Y could plausibly lead to a significant alteration in the probability of missingness, we regard this degree of nonignorability as likely to occur. Therefore, we deem inferences to be sensitive to nonignorability when MinNI is small. Troxel et al. (2004) suggested using a MinNI value of less than 1 as the cutoff for significant sensitivity.

2.5 Sensitivity Analysis Using Probit Selection Model

It might be contended that sensitivity to nonignorability is intrinsically tied to the selection model. Utilizing various binary models for modeling missingness could yield disparate results. Furthermore, there's the risk that the default logistic regression model might be mis-specified, potentially introducing bias. To address this, we employ another prevalent binary model, the Probit model, to compute ISNI and MinNI values.

The probit model maps the linear combination of predictors (i.e., $X\beta$) to a probability between 0 and 1 using a standard normal cumulative distribution function (CDF), specified as $P(G = 1|X) = \Phi(X\beta)$. Consequently, the selection Probit model can be specified as

$$P(G_i = 1|Y_i = y_i, S_i = s_i) = \Phi(\gamma_0 s_i + \gamma_1 y_i)$$

, With the joint log-likelihood function being:

$$\ln L = \sum_{i=1}^{n} \Bigg\{ (1 - g_i) \Big[\ln f_\theta(y_i^{\text{obs}}|x_i) + \ln(1 - \Phi(\gamma_0 s_i + \gamma_1 y_i^{\text{obs}})) \Big] + g_i \ln \int_{\Omega_{y^{\text{mis}}}} f_\theta(y_i^{\text{mis}}|x_i) \Phi(\gamma_0 s_i + \gamma_1 y_i^{\text{mis}}) dy_i^{\text{mis}} \Bigg\},$$

The ISNI derivation for the Probit model (as discussed in section 2.2.2) exhibits minor deviations from that of the Logit model. Notably, $\nabla^2_{\theta,\gamma_1}$, evolves to

$$\nabla^2_{\theta,\gamma_1} = \sum_{i=1}^{n} \left\{ g_i \left[\frac{\phi(\gamma_0 s_i + \gamma_1 y)}{\Phi(\gamma_0 s_i + \gamma_1 y)} \right] \frac{\partial E_\theta(y_i|x_i)}{\partial \theta} \right\}$$

, Here, $\phi(\gamma_0 s_i + \gamma_1 y)$ is the derivative of $\Phi(\gamma_0 s_i + \gamma_1 y)$, representing the PDF of the normal distribution. $\nabla^2_{\theta,\theta}$, the covariance matrix of the observed information, remains unchanged. Thus, the ISNI under the Probit selection model becomes

$$\texttt{ISNI} = -\left[\sum_{i=1}^{n} (1 - g_i) \frac{\partial^2 \ln f_\theta(y_i|x_i)}{\partial \theta^2} \right]^{-1} \sum_{i=1}^{n} \left\{ g_i \left[\frac{\phi(\gamma_0 s_i + \gamma_1 y)}{\Phi(\gamma_0 s_i + \gamma_1 y)} \right] \frac{\partial E_\theta(y_i|x_i)}{\partial \theta} \right\}$$

Both $\phi(.)$ and $\Phi(.)$ can be readily ascertained via the `R` function `dnorm()` and `pnorm()`.

2.6 Interpretation of Probit ISNI and MinNI

Although deriving the ISNI under the Probit model is relatively straightforward, interpreting the nonignorable parameter, γ_1, presents challenges. In the logistic model, a one-unit increase in the

outcome corresponds to a 2.7-fold (i.e., (e^1)) increase in the odds of an observation being missing. However, such direct interpretation is not as evident in the Probit model.

Both the Probit and Logistic models map linear predictors, $g = \gamma_0 s + \gamma_1 y$, to the unit interval. For the Probit model, the probability of being missing is expressed as the standard normal CDF:

$$\begin{aligned} P(G_i = 1) &= \Phi(\gamma_0 s_i + \gamma_1 y_i) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\gamma_0 s_i + \gamma_1 y_i} \exp(-\frac{1}{2} Z^2) dZ \end{aligned}$$

For logistic regression,

$$\begin{aligned} P(G_i = 1) &= \frac{\exp\left(\gamma_0 s_i + \gamma_1 y_i\right)}{1 + \exp(\gamma_0 s_i + \gamma_1 y_i)} \\ &= \int_{-\infty}^{\gamma_0 s_i + \gamma_1 y_i} \frac{\exp(-Z)}{(1 + \exp(-Z))^2} dZ \end{aligned}$$

Here, $f(z) = \frac{\exp(-Z)}{(1+\exp(-Z))^2}$ is the probability density function (PDF) of the logistic distribution. The logistic PDF reveals $E(Z) = 0$ and $Var(Z) = \frac{\pi^2}{3}$. To scale Z to a standardized logistic distribution (with mean 0 and variance 1), we need to modify the density function:

$$f(z) = \frac{\pi}{\sqrt{3}} \frac{\exp(\pi Z/\sqrt{3})}{(1 + \exp(\pi Z/\sqrt{3}))^2}$$

Subsequently, the cumulative distribution function — the logit function — is transformed as:

$$F(z) = \frac{\exp(\pi Z/\sqrt{3})}{1 + \exp(\pi Z/\sqrt{3})}$$

his transformation allows the Logit function to approximate the Probit function (Nelder and Wedderburn, 1972, Fox, 2015)

$$\Phi(\gamma_0 s_i + \gamma_1 y_i) = \frac{\exp(\frac{\pi}{\sqrt{3}}(\gamma_0 s_i + \gamma_1 y_i))}{1 + \exp(\frac{\pi}{\sqrt{3}}(\gamma_0 s_i + \gamma_1 y_i))}$$

Given this transformation, a one-unit increase in Y is associated to a 6.1-fold increase in the odds of an observation being missing ($\exp(\pi/\sqrt{3}) = 6.1$). As suggested by Ma et al. (2005), the MinNI value should be adjusted by a factor of $\pi/\sqrt{3}$,

$$\text{MinNI} = \left| \frac{\pi}{\sqrt{3}} \frac{\sigma_Y \text{SE}}{\text{ISNI}} \right|$$

Such a transformation makes the MinNI values from both the Probit and Logistic Models directly comparable. The various models of sensitivity analysis are detailed further in section 4.4.

Chapter 3

ISNI for Negative Binomial Regression

3.1 Negative Binomial—the Poisson Gamma Mixture Model

Both Poisson and Negative Binomial regressions are frequently employed to model count outcomes. The Poisson distribution includes a single parameter, λ, which acts as both the mean and variance of the count data. Employing Poisson regression is based on the stringent assumption that the mean and variance of the count data are identical.

In practice, the rate λ frequently varies across subjects, leading to observed heterogeneity in count data. This can lead to an increase in the variance of the count data, resulting in overdispersion. Such overdispersion violates the Poisson model's assumption of equidispersion (Mullahy, 1997). To manage this overdispersion, we utilize an alternative distribution for λ that can accommodate overdispersion. A popular approach is to apply a gamma distribution across the mean of the counts (Greene, 2008; Mullahy, 1997). This approach, resulting in a Poisson-Gamma mixture model, gives rise to the negative binomial model.

Consider a Poisson distribution with parameter λ. The probability mass function (PMF) of the Poisson distribution is given by:

$$P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

Suppose that λ is a random variable itself, following a Gamma distribution with shape parameter α and scale parameter $\beta = \frac{p}{1-p}$, where p is the probability of success. The probability density function (PDF) of the Gamma distribution is:

$$P(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$$

The Negative Binomial distribution can be derived as a Poisson-Gamma mixture such that

$$
\begin{aligned}
P(Y=y) &= \int_0^\infty P(Y=y|\lambda)P(\lambda)d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\lambda^{\alpha-1}\frac{e^{-\lambda p/(1-p)}}{\Gamma(\alpha)}(\frac{p}{1-p})^\alpha d\lambda \\
&= \frac{p^\alpha(1-p)^{-\alpha}}{y!\Gamma(\alpha)}\int_0^\infty \lambda^{y+\alpha-1}e^{-\lambda\frac{1}{1-p}}d\lambda \\
&= \frac{p^\alpha(1-p)^{-\alpha}}{y!\Gamma(\alpha)}(1-p)^{\alpha+y}\Gamma(\alpha+y) \\
&= \frac{\Gamma(\alpha+y)}{y!\Gamma(\alpha)}p^\alpha(1-p)^y \\
&= \binom{y+\alpha-1}{y}p^\alpha(1-p)^y,
\end{aligned}
$$

where $Y \sim$ NB (α, p) with mean and variance as

$$
\begin{aligned}
E(Y) &= \frac{\alpha(1-p)}{p} = \mu, \\
Var(Y) &= \frac{\alpha(1-p)}{p^2} = \mu + \frac{1}{\alpha}\mu^2
\end{aligned}
$$

Compared to the Poisson model, the $\frac{\mu^2}{\alpha}$ in the variance captured the data overdispersion.

3.2 Specify the Negative Binomial Regression

In employing the Negative Binomial (NB) distribution to model count data, we can posit that the mean count μ is related to a set of predictors through the link function $\mu = \exp(X'\beta)$. Furthermore, the dispersion parameter is commonly reparameterized as $\frac{1}{\alpha}$, which represents the number of counts required for the $(1/\alpha)^{\text{th}}$ success to occur. This parameterization, often referred to as NB2, offers greater flexibility and is frequently used for parameter estimation(Lord et al., 2012). To distinguish the two pasteurizations, we use κ to denote the dispersion parameter. The resulting NB model is given by:

$$
\begin{aligned}
f(y_i|x_i,\kappa) &= \frac{\Gamma(y+\kappa^{-1})}{y!\Gamma(\kappa^{-1})}\left(\frac{1}{1+\kappa\mu}\right)^{1/\kappa}\left(\frac{\kappa\mu}{1+\kappa\mu}\right)^y \\
&= \frac{\Gamma(y_i+\kappa^{-1})}{y_i!\Gamma(\kappa^{-1})}\left(\frac{1}{1+\kappa\exp(x_i'\beta)}\right)^{1/\kappa}\left(\frac{\kappa\exp(x_i'\beta)}{1+\kappa\exp(x_i'\beta)}\right)^{y_i}
\end{aligned}
$$

The mean and variance of the NB distribution then take on the following forms:

$$E(Y|x) = \mu(x)$$
$$Var(Y|x) = \mu(x) + \kappa\mu(x)^2$$

In this framework, the parameter κ controls the overdispersion. As κ approaches 0, the NB model converges to a Poisson model, thereby bridging the gap between these two commonly used models for count data (Lawless, 1987).

To estimate κ and β_s using the maximum likelihood estimation, the likelihood and log-likelihood functions can be specified as

$$L(\kappa, \beta) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \kappa^{-1})}{y_i!\Gamma(\kappa^{-1})} \left(\frac{1}{1 + \kappa \exp(x_i'\beta)}\right)^{1/\kappa} \left(\frac{\kappa \exp(x_i'\beta)}{1 + \kappa \exp(x_i'\beta)}\right)^{y_i}$$
$$\ln L(\kappa, \beta) = \sum_{i=1}^{n} \Bigg[y_i \ln \kappa + y_i(x_i\beta) - \left(y_i + \frac{1}{\kappa}\right) \ln(1 + \kappa \exp(x_i\beta)) + \ln \Gamma\left(y_i + \frac{1}{\kappa}\right) - \ln y_i! - \ln \Gamma\left(\frac{1}{\kappa}\right) \Bigg]$$

3.3 Specify the Missing Data Model

Let G be an indicator for missing data, where $G_i = 1$ represents missing data for the count outcome y_i, and $G_i = 0$ indicates observed data. We use the same logistic regression to model the probability $P(G_i = 1)$ as introduced in Chapter 2. Therefore, the missing data model can be specified as

$$P(G_i = 1|Y_i = y_i, S_i = s_i) = h(\gamma_0 s_i + \gamma_1 y_i) = \frac{\exp(\gamma_0 s_i + \gamma_1 y_i)}{1 + \exp(\gamma_0 s_i + \gamma_1 y_i)}$$

We then specify the joint likelihood function for the negative binomial model and the missing data model, denoted as $L(\beta, \kappa, \gamma_0, \gamma_1; y^{\text{obs}}, y^{\text{mis}}, g, x, s)$, as follows,

$$L = \prod_{i=1}^{n} \Bigg[\frac{\Gamma(y^{\text{obs}} + \kappa^{-1})}{y^{\text{obs}}!\Gamma(\kappa^{-1})} \left(\frac{1}{1 + \kappa \exp(x_i'\beta)}\right)^{1/\kappa} \left(\frac{\kappa \exp(x_i'\beta)}{1 + \kappa \exp(x_i'\beta)}\right)^{y_i^{\text{obs}}} (1 - h(\gamma_0 s_i + \gamma_1 y_i^{\text{obs}})) \Bigg]^{1-g_i}$$
$$\times \left[\sum_{y_i^{\text{mis}}=0}^{\infty} \frac{\Gamma(y^{\text{mis}} + \kappa^{-1})}{y^{\text{mis}}!\Gamma(\kappa^{-1})} \left(\frac{1}{1 + \kappa \exp(x_i'\beta)}\right)^{1/\kappa} \left(\frac{\kappa \exp(x_i'\beta)}{1 + \kappa \exp(x_i'\beta)}\right)^{y_i^{\text{mis}}} h(\gamma_0 s_i + \gamma_1 y_i^{\text{mis}}) \right]^{g_i}$$

3.4 Derive ISNI for Negative Binomial Model

As indicated in section 2.2.2, the general formula for `ISNI` can be decomposed into two constituent parts: the covariance matrix ($\nabla_{\theta,\theta}$), and the product of the probability of being observed and the derivative of the conditional mean for missing outcome observations (∇_{θ,γ_1}). The conditional mean of the negative binomial regression can be expressed as

$$\mu_i = E(y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip})$$

Therefore, the derivative with respect to β is

$$\nabla_{\theta,\gamma_1} = \frac{\partial E(y_i|x_i)}{\partial \beta} = \exp(\beta^T x_i) x_i$$

While the formula for ∇_{θ,γ_1} in the NB model bears resemblance to that of the Poisson model, the parameter estimates between the two models vary due to their distinct likelihood functions. This leads to different values for ∇_{θ,γ_1}, as shown in Table 2. Moreover, the covariance matrix from the NB ISNI differs from that of the Poisson due to the inclusion of the dispersion parameter κ, which can be challenging to derive. The covariance matrix, the inverse of Fisher's information matrix from the complete data, contains three integral parts: $\frac{\partial^2 \log L(\kappa,\beta)}{\partial \beta_k \partial \beta_l}$, $\frac{\partial^2 \log L(\kappa,\beta)}{\partial \beta_k \partial \kappa}$, and $\frac{\partial^2 \log L(\kappa,\beta)}{\partial \kappa^2}$, $k, l = 1, \cdots, p$ (Lawless, 1987). Specifically,

$$\begin{aligned}
\frac{\partial^2 \log L(\kappa,\beta)}{\partial \beta_k \partial \beta_l} &= -\sum_{i=1}^{n} \frac{(1+\kappa y_i)\exp(x_i\beta) x_{ik} x_{il}}{[1+\kappa \exp(x_i\beta)]^2} \\
\frac{\partial^2 \log L(\kappa,\beta)}{\partial \beta_k \partial \kappa} &= -\sum_{i=1}^{n} \frac{\exp(x_i\beta)\left[y_i - \exp(x_i\beta)\right] x_{ik}}{[1+\kappa \exp(x_i\beta)]^2} \\
\frac{\partial^2 \log L(\kappa,\beta)}{\partial \kappa^2} &= \sum_{i=1}^{n} \left\{ \psi'(y_i + \frac{1}{\kappa}) - \psi'(\frac{1}{\kappa}) - \frac{y_i}{\kappa^2} - \frac{2\ln[1+\kappa\exp(x_i\beta)]}{\kappa^3} + \right. \\
&\qquad \left. \frac{2\exp(x_i\beta)}{\kappa^2[1+\kappa\exp(x_i\beta)]} + \frac{\kappa\exp(x_i\beta)(y_i+1)}{\kappa[(1+\kappa\exp(x_i\beta)]^2} \right\}
\end{aligned}$$

$\psi(z) = \frac{d}{dz}\log\Gamma(z)$ is the digamma function, and its derivative, $\psi'(z)$, can be conveniently evaluated using the readily `trigamma` function in R.

The frequently used R function `glm.nb()` provides only the expected information matrix. As Lawless (1987) noted, the maximum likelihood estimators κ and β_s are asymptotically uncorrelated as $n \to \infty$ (i.e., $\frac{\partial^2 \log L(\kappa,\beta)}{\partial \beta_k \partial \kappa} = 0$). However, when the sample size is not large, the assumption of asymptotic behavior may not be accurate.

In response to this issue, we incorporated the calculation of the observed information matrix in our newly developed `isniglm.nb()` R function. This allows us to compute the NB ISNI index,

thus providing more accurate results when dealing with limited sample sizes. The `isniglm.nb()` R function will be added to the `isni` R package.

3.5 ISNI Comparison between Negative Binomial and Poisson Models–Simulation Study

3.5.1 Simulation 1: Comparing the ISNI index for Negative Binomial Model to Poisson Model

To investigate the performance of the NB ISNI method and the newly developed R package function `isniglm.nb()`, we conducted a simulation study in which the dataset exhibited nonignorable missingness. This dataset comprised 500 observations along with one covariate, X_1, which was generated from a normal distribution with mean 0 and standard deviation 1. The mean counts, denoted by μ, were determined using the equation:

$$\mu = \exp(1 + 0.5 * X_1) \tag{3.1}$$

The dispersion parameter α was fixed at 2. Subsequently, the count outcome y was generated using the parameters μ and α under a negative binomial distribution, operated as `y=rnbinom`(N, Size=α, mu=μ). Note that the current R negative binomial functions parameterize dispersion as α; hence, a smaller α indicates a higher degree of dispersion. The simulated count outcome has a mean of 3.014, with a range of 0 to 29.

To simulate missingness in the outcome, we calculated the missing probability, $\mathrm{P_{mis}}$, for each simulated outcome y. As our objective was to simulate nonignorable missingness, this probability was inherently tied to the outcome y with coefficient 0.5 and covariate X_1 using a logistic function, as given by:

$$\mathrm{P}(\mathrm{G}_i = 1|x_i, y_i) = \frac{\exp(-2 + 2 * X_{1i} + 0.5 * y_i)}{1 + \exp(-2 + 2 * X_{1i} + 0.5 * y_i)} \tag{3.2}$$

Utilizing $\mathrm{P_{mis}}$, we simulated the missing indicator G using the binomial distribution, where $\mathrm{G} = 1$ indicates that the outcome y is missing. In this single simulation case, approximately 41.2% of the outcomes were set to be missing.

Table 1: *ISNI Simulated Results for NB and Poisson Models*

	Full NB			MAR NB				MAR Poisson			
Parameter	Estimate	SE	p-value	MAR est.	SE	ISNI	MinNI	MAR est.	SE	ISNI	MinNI
Intercept	0.962	0.043	$< 2e^{-16}$	0.493	0.070	0.667	0.167	0.484	0.054	0.362	0.235
X_1	0.480	0.042	$< 2e^{-16}$	0.143	0.075	0.511	0.232	0.126	0.058	0.254	0.362
α	1.987	0.242	-	2.277	0.534	-0.119	7.101	-	-	-	-

As shown in Table 1, the estimated parameters derived from the complete dataset closely match the original parameters—specifically, Intercept = 1, $\beta = 0.5$, $\alpha = 2$—as outlined in equation 3.1. In the MAR NB model, the estimates for intercept and X_1 were significantly lower than expected, while the estimated α was higher. Additionally, the standard errors for Intercept, X_1, and α were greater than those in the complete NB model. This underscores the possible impact on parameter estimates due to the presence of missing data. Moreover, the MinNI indices for both the intercept and X_1 fell below 1, indicating a high degree of sensitivity to nonignorable missingness.

Upon running the Poisson ISNI analysis on the identical simulated dataset, the Poisson MAR estimates yielded parameter estimates considerably similar to those from the MAR NB model. The MinNI indices for both the intercept and X_1 fell below 1, reaffirming the sensitivity to the nonignorability observed in the MAR NB model. However, a notable disparity was evident in the ISNI values between the MAR models, primarily due to differences in the covariance matrix ($-\nabla_{\theta,\theta}^{-1}$). Table 2 illustrates a comparison of covariance matrices between the NB (columns 1-3) and Poisson (columns 5-6) models. The variances of the intercept and β in the MAR NB model were 1.8 and 1.9 times larger, respectively, than those in the MAR Poisson model. Moreover, the covariance between the intercept and β was twice as high in the NB model. Consequently, this led to a larger ISNI value in the NB model relative to the Poisson model, given the similar estimates of ∇_{θ,γ_1}.

Table 2: Comparison for $-\nabla_{\theta,\theta}^{-1}$ and ∇_{θ,γ_1} between NB and Poisson Models

	NB model				Poisson Model		
Parameter	Intercept	β	α	∇_{θ,γ_1}	Intercept	β	∇_{θ,γ_1}
Intercept	0.005163	0.003019	-0.000706	113.77945	0.002886	0.001494	112.39327
X_1	0.003019	0.006368	-0.001463	26.32379	0.001494	0.003388	25.41806
α	-0.000706	-0.001463	0.285161	0	-	-	-

3.5.2 Simulation 2: Investigating the Impact of Overdispersion Degree on ISNI

Dispersion plays a crucial role in modeling count data. However, the influence of varying degrees of dispersion on the sensitivity of parameter estimates is not entirely understood. As noted in Section 3.1, as the parameter α increases, the variance of the outcome progressively approaches the mean. Subsequently, overdispersion diminishes towards equal dispersion - a vital assumption for employing the Poisson model.

To investigate this, we simulated 1,000 datasets, each defined by the coefficient parameters outlined in Equation 3.1, across a range of α values. We systematically adjusted the α parameter from 1 to 100, resulting in a total of 100,000 (1,000 $\times$ 100) datasets. As α increases, we observe a decreasing degree of overdispersion. For each dataset, we conducted ISNI analyses for both

Negative Binomial and Poisson models. Subsequently, we plotted the mean ISNI and MinNI values with their 95% confidence intervals, against each α value. These plots were generated for both the intercept and the X_1 variable, providing a comprehensive visual representation of the trends across varying levels of α.

Dispersion is pivotal in evaluating the sensitivity of nonignorable missing data. As shown in Figure 2, strong dispersion—indicated by smaller α values—leads to higher estimated ISNI values and smaller MinNI values in the Negative Binomial model. This suggests an increased degree of sensitivity to nonignorability. Conversely, as the degree of dispersion diminishes (signified by larger α values), the estimated ISNI values decrease and the MinNI values increase, pointing to reduced sensitivity.

Interestingly, the ISNI value appears relatively stable across different levels of dispersion in the Poisson model. This highlights a key limitation of the Poisson model when dealing with varying dispersion conditions: it tends to underestimate parameter standard errors and can distort statistical significance in the presence of overdispersion. As a result, caution should be exercised when using the Poisson model for overdispersed data.

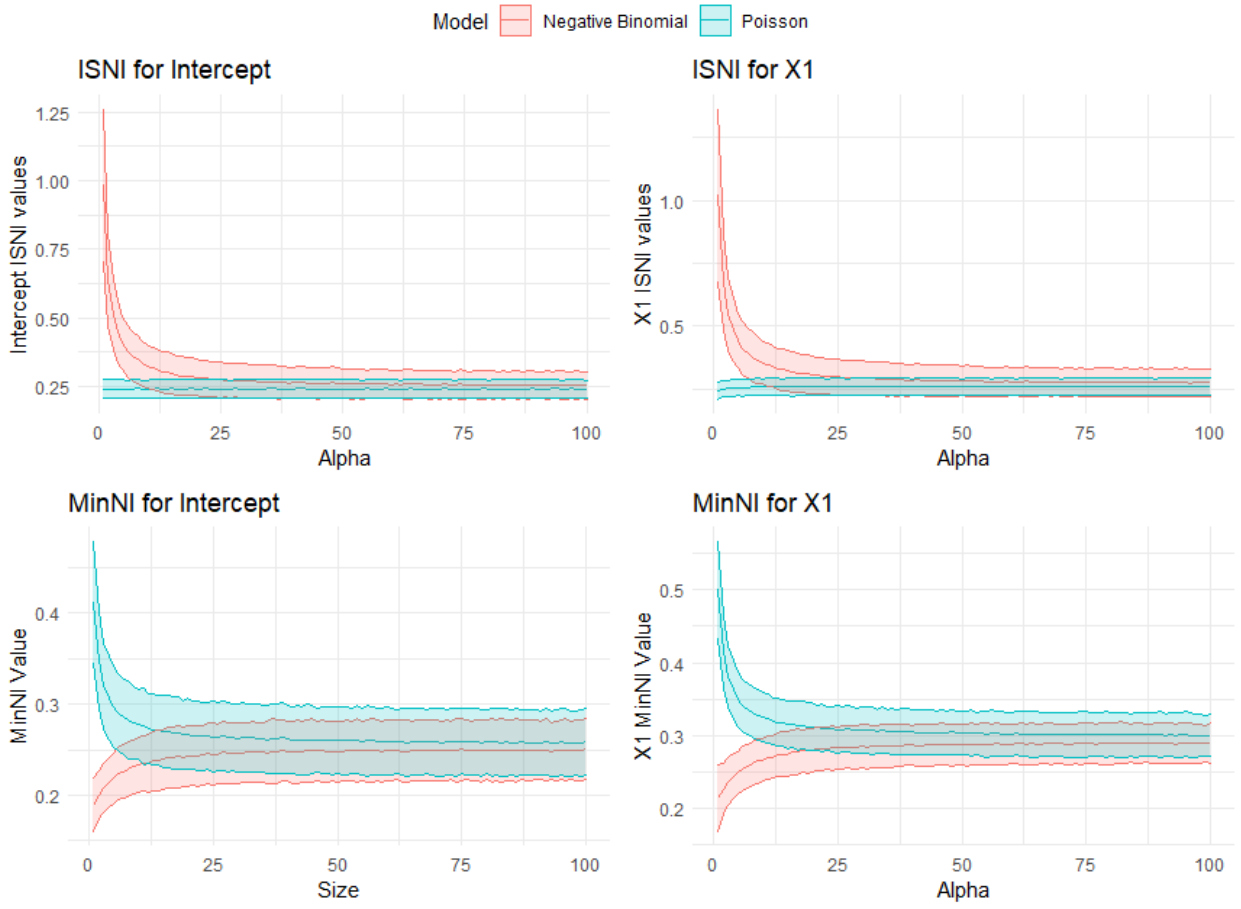
3.5.3 Simulation 3: Examining the Impact of Missing Data Proportion on ISNI

Simulation 2 unveils the influence of dispersion on ISNI estimates. Interestingly, the proportion of missing data can modulate the sensitivity to nonignorability. For instance, in the univariate normal scenario, the ISNI value directly hinges on the proportion of missing data (n_m/n), as delineated by Troxel (2004). However, studies exploring the relationship between the proportion of missing data and sensitivity to nonignorability are still lacking.

Maintaining the same parameter setting, we generated eight datasets where the proportion of missing data ranged from 0.1 to 0.8, incrementing by 0.1. We kept the dispersion variable at 2. Different intercept values in Equation 3.2 were used to control the proportion of missing data. Figure 3 illustrates the influence of various missing proportions on both the mean and variance. Notably, the variance was significantly impacted by the missing data, with the blue dashed line significantly below the blue solid line. Both the mean and variance decrease in response to an increasing proportion of missing data.

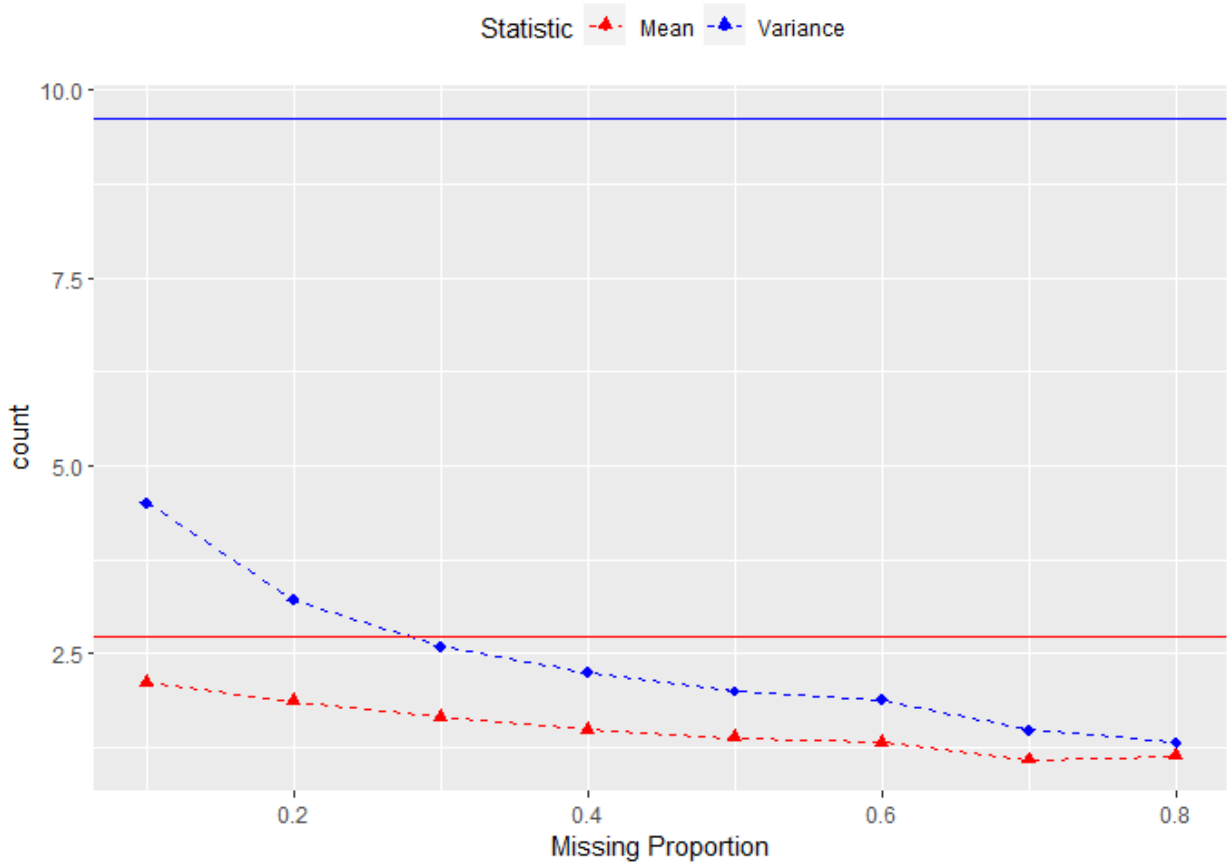
Subsequently, we generated 1000 datasets for each missing proportion and computed the mean and 95% confidence intervals of the ISNI and MinNI values for both the intercept and the coefficient β. Figure 4 illustrates that a greater proportion of missing data results in higher ISNI values and lower MinNI values for both the intercept and β in the Negative Binomial model. In particular, the association between the ISNI values and the missing proportion demonstrates a monotonic trend. This finding further implies that an increased missing proportion correlates with a higher level of sensitivity to nonignorability. The trend observed in the NB model is reaffirmed by the ISNI

Figure 2: ISNI and MinNI estimates Comparison at Different Dispersion Levels



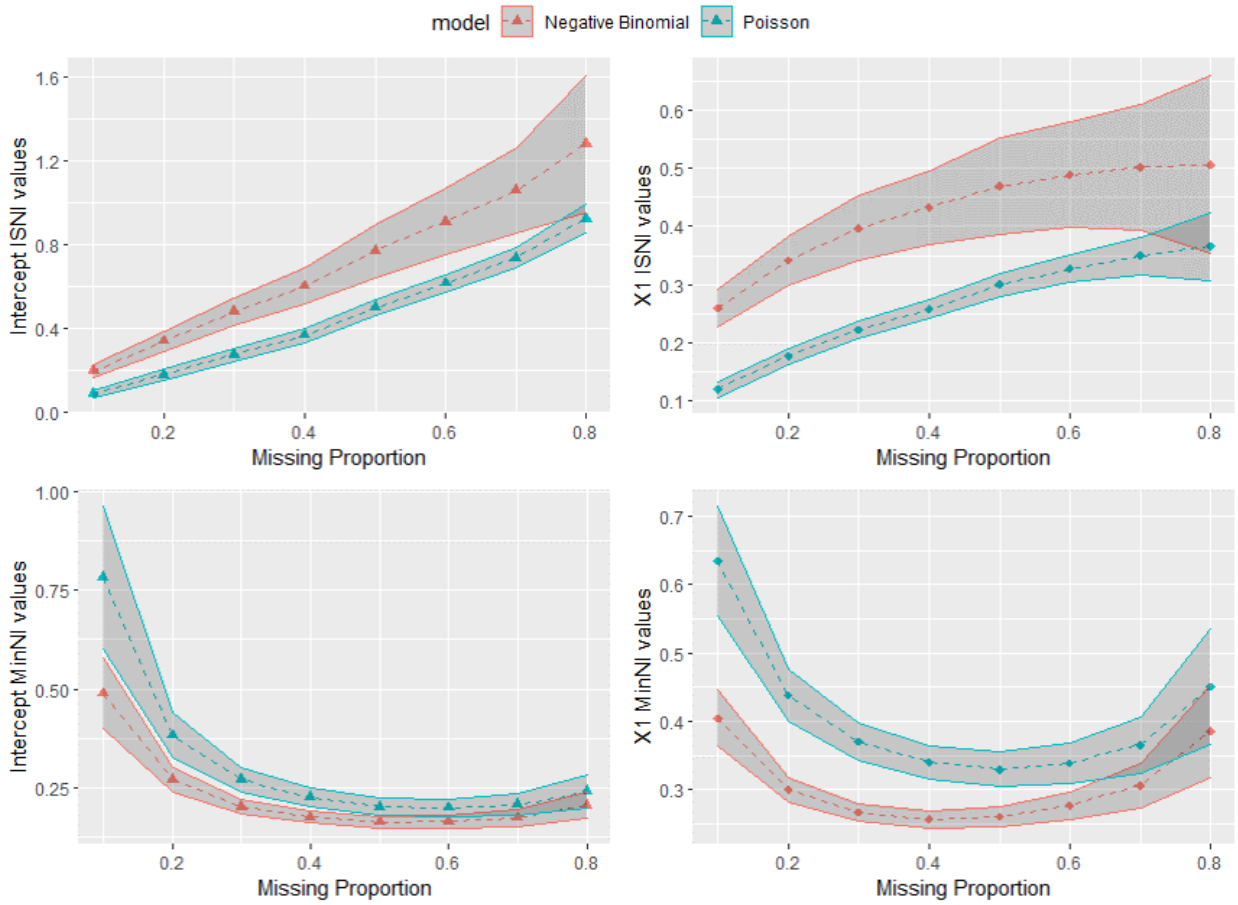
values derived from the Poisson model, although the latter is lower. This difference in ISNI values between the Poisson and NB models is attributed to the dispersion, which was elaborated on in Section 3.5.2. On the other hand, the MinNI values for β exhibit a quadratic trend, attributable to the rising estimated standard error of β as the proportion of missing data increases.

Figure 3: Mean and Variance under Various Missing Proportion



Note: The solid red and blue lines represent the mean and variance of the full data. The dashed red and blue lines represent the mean and variance of the complete data

Figure 4: ISNI estimates Comparison at Various Missing Proportion



Chapter 4

Sensitivy Analysis for Activities of Daily Living (ADL) following Cardiac Surgery

4.1 Background

Coronary artery bypass graft (CABG) surgery, a prevalent surgical treatment for coronary heart disease among elderly patients, reroutes the blood pathway to bypass narrowed or clogged arteries, enhancing blood flow and oxygen supply to the heart (Mori et al., 2021). Despite its invasive nature, it is deemed safer with lower mortality rates (Diodato and Chedrawy, 2014). Alternatively, percutaneous coronary intervention (PCI) presents a less invasive procedure to CABG for coronary treatments. Clinical trials suggest that PCI is not inferior to CABG in addressing primary coronary artery disease (Serruys et al., 2009).

The ADL encompasses six functional tasks: bathing, dressing, eating, toileting, walking across a room, and getting in and out of bed (Katz, 1963. The ADL score indicates the number of difficulties experienced in performing these tasks, with scores ranging from 0 (no difficulties) to 6 (difficulties in all tasks). ADL is a critical indicator of elders' daily functionality. Issues concerning ADLs often signify potential health and cognitive impairments. The deterioration of ADL capacities not only suggests functional impairment but also signals an elevated level of frailty. In this study, we used the ADL score collected through the Health and Retirement Study (HRS) for aging research. It is common for elderly individuals discharged from hospitalization to experience worse ADL functioning than upon admission, making them at high risk for poor functional outcomes due to the challenge of regaining pre-admission levels and the likelihood of new functional deficits occurring during or post-hospitalization (Covinsky et al., 2003).

Although both Coronary Artery Bypass Grafting (CABG) and Percutaneous Coronary Intervention (PCI) procedures effectively address blocked cardiac arteries, it remains uncertain which method might exacerbate Activities of Daily Living (ADLs). As such, the primary objective of

this study is to compare the postoperative impact of these treatments on Activities of Daily Living (ADLs) in elderly patients. We hypothesize that CABG may result in a higher level of ADL impairment in comparison to the PCI procedure due to the invasive nature of the procedure.

Moreover, a considerable number of post-surgery ADL measurements are missing. If the mechanism causing this missingness is not at random and is influenced by unobserved outcome values, it could lead to biased estimation of treatment effects. To mitigate this, we plan to conduct a sensitivity analysis on the missing outcome using the ISNI method.

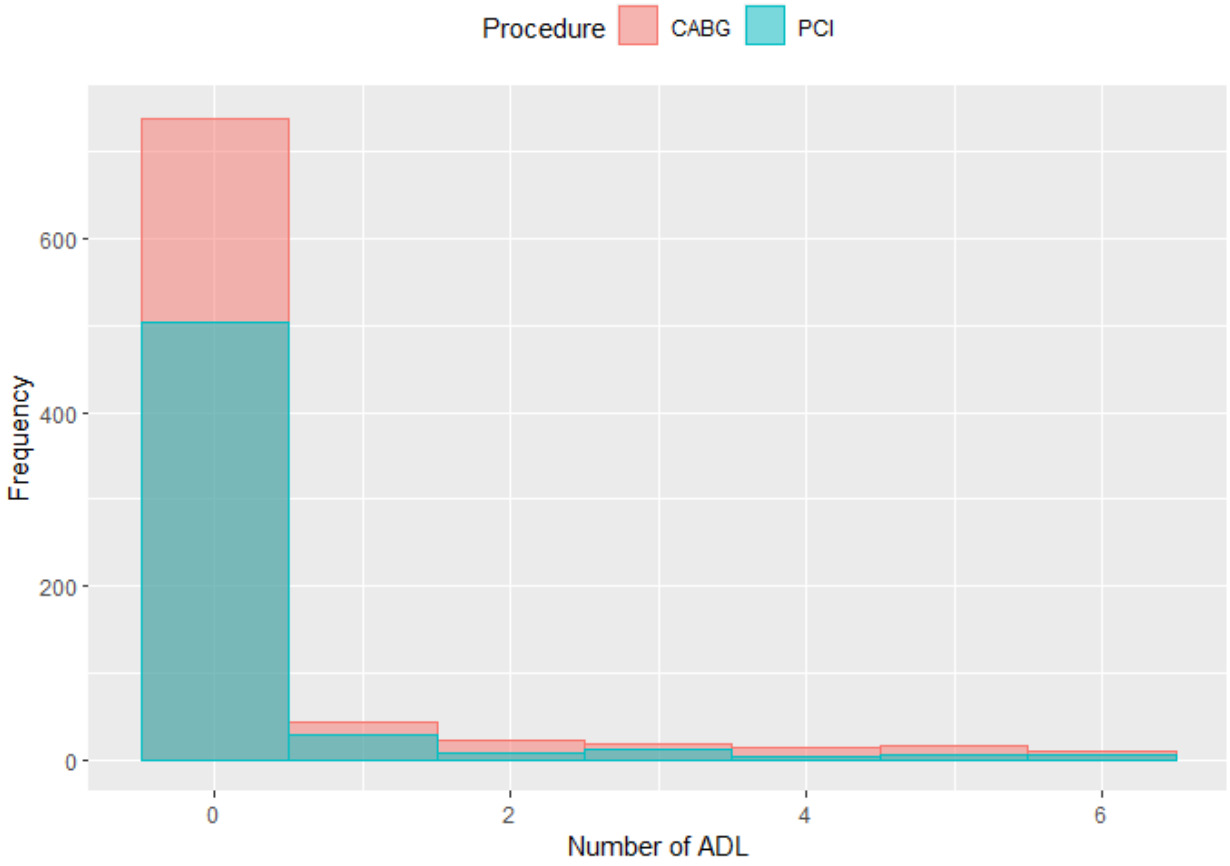
4.2 Cohort Description

The ADL scale ranges from 0, indicating no difficulties, to 6, signifying difficulties in all six measurements. Although the upper bound theoretically does not extend to infinity, in clinical and epidemiological contexts, ADLs are often modeled through count models (Zaninotto and Falaschetti, 2011). In a recent HRS study, a comparison for memory loss was drawn between the CABG and PCI procedures (Whitlock et al., 2021). Here, In this study, we used the same cohort to compare the decline in post-surgical ADLs. Of the subjects, 1015 underwent PCI procedures and 665 had CABG surgeries, with a mean age of 74 (SD, 6.4), 59.6% being male, and 83.4% being white (Table 3). The baseline ADL score, taken before the cardiac procedures, showed no significant difference between the two groups (p-value=0.4501). We also gathered other socioeconomic variables and self-reported comorbidities, which were listed in Table 3. Figure 5 illustrated a comparison of postoperative ADLs between the CABG and PCI procedures. It indicated that the majority of patients did not suffer any postoperative function difficulties. However, about 14.8% of postoperative ADLs were not reported. This could be a cause for concern as the similar distribution of ADLs between the two procedures might be an artifact of nonignorable missingness, potentially misrepresenting the true distribution within the underlying cohort.

Listing 4.1: HRS ADL data display

```
## Display the first six rows of the HRS ADL data
> head(ADL_cohort)
ID Procedure ADL age gender diabetes falls hypertension heartfailure pain stroke marriage
1         1   NA  65      1        0     0            1           0     0      0        1
2         0   0   67      1        0     1            0           0     1      1        1
3         1   0   70      1        1     0            1           0     0      0        1
4         0   NA  67      0        1     1            0           0     0      0        2
5         0   NA  79      1        0     0            1           0     1      0        1
6         0   0   66      1        0     0            0           0     0      0        2
```

Figure 5: ADL Histogram Comparison between CABG and PCI



4.3 Modeling ADL and Conducting ISNI Sensitivity Analysis

A demonstration of the HRS ADL data is displayed in Listing 4.1. Our initial approach was to use Poisson regression to model the ADL outcome. However, the overdispersion test, which evaluates $\kappa = 0$ in $\text{var}(y_i) = \mu + \kappa g(\mu)$, indicated strong evidence of overdispersion (p-value=1.494e-11) (Cameron and Trivedi, 1990). As a result, we chose to use negative binomial regression. The conditional mean of ADL for NB regression was as outlined below:

$$\begin{aligned} \text{E}(\text{ADL}_i) = \mu_i = \exp(&\beta_0 + \beta_1 \text{procedure}_i + \beta_2 \text{age}_i + \beta_3 \text{gender}_i + \beta_4 \text{diabetes}_i + \beta_5 \text{falls}_i \\ &+ \beta_6 \text{hypertension}_i + \beta_7 \text{heartfailure}_i + \beta_8 \text{pain}_i + \beta_9 \text{stroke}_i + \beta_{10} \text{married}_i), \end{aligned} \tag{4.1}$$

β_1 in equation 4.1 signifies the log difference in ADL counts between the PCI and CABG procedures, with CABG serving as the reference group. To account for potential confounding factors, we incorporated variables such as age, gender, marital status, and comorbidities into the model. The results under MAR are presented in the first four columns of Table 4.

Under MAR, the procedure type is not identified as a significant factor of ADLs. There is no significant difference in ADLs between the PCI and CABG procedures. Age, diabetes, heart failure, pain, and stroke are all statistically significant risk factors for higher ADL levels, as shown in Table 4. On average, a patient's ADL score worsens by approximately 1.03 with each passing year. Patients diagnosed with diabetes, heart failure, pain, or those who have experienced stroke episodes tend to exhibit higher ADL levels, on average increasing by approximately 2.05, 2.32, 1.63, and 2.9 respectively.

We subsequently conduct the ISNI analysis using the method developed in Chapter 3 for the moderate missing ADLs. The missing model is specified as follows

$$\begin{aligned}\Pr(\mathtt{G}_i = 1) = \text{logit}[&(\text{procedure}_i, \text{age}_i, \text{gender}_i, \text{diabetes}_i, \text{falls}_i, \text{hypertension}_i, \\ &\text{heartfailure}_i, \text{pain}_i, \text{stroke}_i, \text{marriage}_i)^T \gamma_0 + \gamma_1 * \mathtt{ADL}_i)]\end{aligned} \tag{4.2}$$

Listing 4.2: ISNI for NB and Poisson Model

```
## Specify the Y model
ymodel.adl<-ADL~as.factor(procedure)+age+as.factor(gender)+
             as.factor(diabetes)+as.factor(falls)+
             as.factor(hypertension)+as.factor(heartfailure)+
             as.factor(pain)+as.factor(stroke)+as.factor(marriage)
## running the ISNI for NB model from the isniglm.nb() function
fit.adl<-isniglm.nb(ymodel.adl,data=ADL_cohort)
summary(fit.adl)
## running the ISNI for Poisson model from the isniglm() function
fit.adl.pos<-isniglm(ymodel.adl,data=ADL_cohort,family="poisson")
summary(fit.adl.pos)
```

The detailed implementation and codes of ISNI for NB model was listed in listing 4.2. The ISNI results for the NB model are shown in columns 4-5 of Table 4. Although the type of procedure is not statistically significant, it exhibits high sensitivity to nonignorable missingness, as indicated by a MinNI score of 0.622. Given that the estimate for the procedure is negative, the difference in Activities of Daily Living (ADLs) between PCI and CABG could be further adjusted downward if $\gamma_1 > 0$, shifting from $\exp(-0.306)$ to $\exp(-0.306 - 0.334)$. Factors such as age, gender, diabetes, falls, heart failure, pain, and stroke are all sensitive to nonignorable missingness. The ISNI values for heart failure and stroke are notably high. Considering that the effects under the MAR assumption were positive, the estimated effects for heart failure and stroke could be further adjusted upward if $\gamma_1 > 0$.

We also implemented the ISNI analysis under the Poisson model (Table 4, columns 6-9). While the MAR parameter estimates exhibited similarities across the two models, significant disparities were noticed in the standard error estimates, ISNI values, and MinNI indices. Notably, none of the covariates in the Poisson model were sensitive to nonignorable missingness, standing in stark contrast to the results from the Negative Binomial model. As highlighted in the simulation study

in Section 3.5.2, a smaller α value tends to widen the ISNI difference between the NB and Poisson models. In this specific case, the estimated α value is 0.127, which contributes to a larger dispersion in the variance estimate, leading to a substantial difference in ISNI estimation. This further suggests that the Poisson model, given its large dispersion, may not be appropriate for use in this scenario.

4.4 Sensitivity Analysis Using Different Selection Models

In addition to the ISNI and MinNI values detailed in Table 4, we conducted a sensitivity analysis using the Probit model. Furthermore, we adjusted for various predictors in the selection model, employing both logistic and probit approaches. The models evaluated include the full model, a backward-selected model incorporating age, diabetes, falls, heart failure, stroke, and marital status, as well as an intercept-only model.

Table 5 presents the ISNI and MinNI values for both Logistic and Probit selection models. For the Backwards-selected and full models, ISNI values were similar within each selection model, excluding the intercept-only model. The distinction between selection models (Logistic vs. Probit) arises primarily from ∇_{θ,γ_1}. Conversely, the MinNI values for the two selection models were closely aligned, suggesting that the transformation by the scaling factor $\pi/\sqrt{3}$ renders their sensitivities comparable.

Furthermore, the choice of different models and selection models did not change the sensitivity direction to nonignorability. Specifically, the factors `Intercept`, `procedure`, `age`, `gender`, `diabetes`, `falls`, `heart failure`, `stroke`, and `marital status` consistently demonstrated sensitivity to nonignorable missingness. This was evidenced by the MinNI values under various models consistently falling below 1. It underscores that the choice of selection model does not alter the sensitivity to nonignorability.

Table 3: Risk Factors Summary

Factors	PCI (N=1015)	GABG (N=665)	Overall (N=1680)	P-value
age (mean,SD)	74.0(6.60)	73.7(6.02)	73.9(6.38)	0.6765
gender(male)	556(54.8%)	446(67.1%)	1002(59.6%)	<0.0001
Race				0.0064
White	829(81.7%)	572(86%)	1401(83.4%)	
Black	98(9.7%)	41 (6.2%)	139 (8.3%)	
Hispanic	61(6%)	45(6.8%)	106 (6.3%)	
Other	27(2.7%)	7(1.1%)	34(2%)	
ADL at baseline (mean, SD)	0.2(0.65)	0.1(0.57)	0.1(0.62)	0.4501
Education				0.0369
0. <high school/GED	329(32.4%)	220(33.1%)	549(32.7%)	
1. high school	343(33.8%)	200(30.1%)	543(32.3%)	
2. some college	192(18.9%)	113(17%)	305(18.2%)	
3. college and above	151(14.9%)	132(19.8%)	283(16.8%)	
Married or partnered	368(36.3%)	185(27.8%)	553 (32.9%)	0.0003
Self-reported Comorbidities				
Diabetes	314(30.9%)	211(31.7%)	525(31.3%)	0.7315
Falls	355(35%)	193(29%)	548(32.6%)	0.0109
Hypertension	676(66.6%)	457(68.7%)	1133(67.4%)	0.3643
Heart Failure	63(6.2%)	36(5.4%)	99(5.9%)	0.4995
Pain	302(29.8%)	173(26%)	475(28.3%)	0.0961
Stroke	147(14.5%)	88(13.2%)	235(14%)	0.4702

Table 4: NB and Poisson MAR Models and ISNI Results

	NB model					Poisson Model			
Parameter	MAR Est.	Std. Err	Pr(>|z|)	ISNI	MinNI	MAR Est.	Std. Err	ISNI	MinNI
Intercept (β_0)	-4.235	1.128	0.000	-4.598	0.263	-3.737	0.538	-0.430	1.343
procedure (PCI vs.CABG) (β_1)	-0.306	0.194	0.114	-0.334	0.622	-0.276	0.099	0.005	23.436
age (β_2)	0.032	0.015	0.031	0.062	0.259	0.026	0.007	0.007	1.092
gender:male (β_3)	-0.261	0.212	0.218	-0.248	0.918	-0.220	0.106	0.002	54.877
diabetes (β_4)	0.720	0.200	0.000	1.043	0.206	0.583	0.095	0.049	2.065
falls (β_5)	0.301	0.201	0.134	0.487	0.443	0.366	0.095	0.078	1.305
hypertension (β_6)	0.175	0.208	0.400	0.150	1.484	0.210	0.112	0.033	3.712
heart failure (β_7)	0.842	0.369	0.022	1.980	0.200	0.886	0.126	0.083	1.623
pain (β_8)	0.491	0.205	0.016	0.211	1.043	0.427	0.096	-0.049	2.117
stroke (β_9)	1.067	0.251	0.000	2.021	0.133	0.797	0.101	0.075	1.447
marriage (β_{10})	0.059	0.220	0.787	-0.091	2.589	0.145	0.107	0.008	14.253
alpha	0.127	0.015		-0.004	4.137				

Table 5: Selection Models ISNI and MinNI comparison

				Logistic Model						Probit Model					
				Full		Backwards-selected		Intercept-only		Full		Backwards-selected		Intercept-only	
Parameter	MAR Est.	Std. Err	Pr(>|z|)	ISNI	MinNI	ISNI	MinNI	ISNI	MinNI	ISNI	MinNI	ISNI	MinNI	ISNI	MinNI
Intercept (β_0)	-4.235	1.128	0.000	-4.598	0.263	-4.621	0.263	-6.333	0.191	-7.390	0.297	-7.409	0.296	-11.602	0.189
procedure (PCI vs.CABG) (β_1)	-0.306	0.194	0.114	-0.334	0.622	-0.320	0.650	-0.367	0.567	-0.582	0.648	-0.539	0.699	-0.672	0.561
age (β_2)	0.032	0.015	0.031	0.062	0.259	0.062	0.258	0.085	0.189	0.100	0.290	0.101	0.288	0.155	0.187
gender:male (β_3)	-0.261	0.212	0.218	-0.248	0.918	-0.271	0.837	-0.393	0.579	-0.412	0.999	-0.467	0.883	-0.720	0.573
diabetes (β_4)	0.720	0.200	0.000	1.043	0.206	1.039	0.207	1.294	0.166	1.759	0.221	1.744	0.223	2.372	0.164
falls (β_5)	0.301	0.201	0.134	0.487	0.443	0.491	0.439	0.627	0.345	0.794	0.492	0.802	0.487	1.148	0.340
hypertension (β_6)	0.175	0.208	0.400	0.150	1.484	0.164	1.366	0.182	1.225	0.251	1.612	0.285	1.421	0.334	1.212
heart failure (β_7)	0.842	0.369	0.022	1.980	0.200	2.002	0.198	2.717	0.146	3.277	0.219	3.318	0.216	4.977	0.144
pain (β_8)	0.491	0.205	0.016	0.211	1.043	0.193	1.140	0.197	1.114	0.375	1.061	0.342	1.165	0.361	1.102
stroke (β_9)	1.067	0.251	0.000	2.021	0.133	2.029	0.133	2.595	0.104	3.381	0.144	3.382	0.144	4.754	0.103
marriage (β_{10})	0.059	0.220	0.787	-0.091	2.589	-0.103	2.298	-0.171	1.377	-0.155	2.765	-0.182	2.353	-0.314	1.363
alpha	0.127	0.015		-0.004	4.137	-0.004	4.107	-0.005	3.126	-0.006	4.518	-0.006	4.493	-0.009	3.095

Chapter 5

Discussion

This thesis advanced the ISNI index for the negative binomial regression model when modeling count outcomes with missing data. Building on the previous work of Troxel et al., 2004, we further derived the observed covariance matrix and ∇_{θ,γ_1} for the NB ISNI formula. To fulfill this purpose, the R function isniglm.nb() was developed (see Appendix A). Subsequent simulation studies were performed to demonstrate the application of this R function. We undertook a comparison of ISNI results, MinNI, the covariance matrix, and ∇_{θ,γ_1} between the Negative Binomial and Poisson Models. Additionally, we explored ISNI results across various degrees of dispersion. Lastly, we employed the NB ISNI method in a real-world scenario, investigating the potential impact of different cardiac surgery procedures on patients' post-operative activities of daily living (ADL). By implementing both the NB and Poisson models in a real-world context, the ISNI outcomes corroborated the findings from the simulation study.

The dispersion parameter α holds considerable sway over both the estimation of the variance and the sensitivity to nonignorability. In our parameterization scheme, where $\text{Var(Y)} = \mu + \frac{\mu^2}{\alpha}$, a smaller α represents a larger degree of overdispersion. In addition, α serves as a key determinant in the derivation of Fisher's information, as it is incorporated into the variance and covariance estimation during the computation of the second derivative (as shown in section 3.4). This implies that the constituent elements of the covariance matrix can vary by various levels of α. For instance, a diminished α will enlarge the elements housed within the covariance matrix. This enlargement has a direct impact on the ISNI values; a lower α tends to result in a higher ISNI value. This, in turn, cultivates a higher degree of sensitivity to nonignorability, reflected in the reduced MinNI values.

The findings from Simulation 2 reaffirmed the impact of dispersion. As α rises from 1 to 100, there is a corresponding decrease in the ISNI values for both the intercept and β. This, in turn, leads to a reduced sensitivity to nonignorability as dispersion diminishes. In the end, as α increases, the ISNI values of the NB and Poisson Models tend to converge. An intriguing observation from Simulation 3 is the influence of the missing proportion on the degree of sensitivity to nonignorability.

Specifically, when data is not missing at random, there is a monotonic relationship between the missing proportion and ISNI values.

The routine practice for modeling count outcomes often involves testing for overdispersion. This becomes more critical when data is missing because different levels of dispersion alter the sensitivity to nonignorable missingness. More specifically, a higher degree of overdispersion leads to larger ISNI values, thereby intensifying the sensitivity to nonignorability. In the ADLs example, while the parameter estimates remained similar between the NB and Poisson models, the standard errors exhibited significant divergence due to the overdispersion. Furthermore, the NB MAR estimates were notably subject to nonignorable missingness, despite the moderate percentage of missing data (14.8%).

The ISNI and MinNI results were consistent across different selection models. While the calculation of ∇_{θ,γ_1} varies between the Logistic and Probit models, the MinNI values became comparable after the appropriate scaling transformation. The number of predictors used in the selection models did not influence the sensitivity levels. Due to the straightforward interpretation of logistic regression, we opted to incorporate the logistic model as the default selection model in the newly developed `isniglm.nb()` R function.

Our research should be evaluated in light of strengths and limitations. To the best of our knowledge, this is the first method of sensitivity analysis specifically designed for overdispersed count data. Additionally, this study advanced the ISNI index for missing count data, highlighting the significant impact of dispersion on the sensitivity to nonignorability. The method serves as a valuable tool for researchers, aiding in the selection of appropriate models and the execution of precise sensitivity analyses in the presence of data overdispersion. On the other hand, a limitation of the `isni` package is its current focus on evaluating sensitivity to nonignorability solely for missing values in the outcome variable. Missing values in covariates must be imputed before running ISNI analyses (e.g., Woodward et al., 2021). While methods for assessing sensitivity to nonignorable missingness in both outcomes and covariates do exist (Gao et al., 2016; Yuan et al., 2020), these approaches typically involve joint modeling of covariates, outcomes, and MDMs. One avenue for future research would be to extend the `isni` package to handle nonignorable missingness in both outcomes and covariates. Moreover, while the Negative Binomial model can accommodate overdispersed count data, it is less effective than zero-inflated models in handling outcomes with excessive zeros. Therefore, another promising avenue for future research would be to develop ISNI indices specifically for zero-inflated models.

In summary, we have developed the ISNI index for the negative binomial model. The R function, isniglm.nb(), was created and is readily available for the implementation of this method. Our investigations revealed that, compared to the Poisson model, the sensitivity of parameters to nonignorability is significantly affected by the degree of dispersion. It is our hope that this method will assist researchers in employing the correct model for count outcomes in the presence of overdispersion, and in conducting accurate sensitivity analyses when data is missing.

Bibliography

Bell, M. L., Fiero, M., Horton, N. J., & Hsu, C.-H. (2014). Handling missing data in rcts; a review of the top medical journals. *BMC medical research methodology*, *14*(1), 1–8.

Blankers, M., Koeter, M. W., Schippers, G. M., et al. (2010). Missing data approaches in ehealth research: Simulation study and a tutorial for nonmathematically inclined researchers. *Journal of Medical Internet Research*, *12*, e1448.

Brown, R. T., Diaz-Ramirez, L. G., Boscardin, W. J., Lee, S. J., & Steinman, M. A. (2017). Functional impairment and decline in middle age: A cohort study. *Annals of internal medicine*, *167*(11), 761–768.

Brown, R. T., Diaz-Ramirez, L. G., Boscardin, W. J., Lee, S. J., Williams, B. A., & Steinman, M. A. (2019). Association of functional impairment in middle age with hospitalization, nursing home admission, and death. *JAMA internal medicine*, *179*(5), 668–675.

Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the poisson model. *Journal of econometrics*, *46*(3), 347–364.

Catherine, N. L. A., Boyle, M., Zheng, Y., McCandless, L., Xie, H., Lever, R., Sheehan, D., Gonzalez, A., Jack, S. M., Gafni, A., et al. (2020). Nurse home visiting and prenatal substance use in a socioeconomically disadvantaged population in british columbia: Analysis of prenatal secondary outcomes in an ongoing randomized controlled trial. *Canadian Medical Association Open Access Journal*, *8*, E667–E675.

Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. *Psychological Methods*, *21*, 427.

Chen, H. Y., Xie, H., & Qian, Y. (2011). Multiple imputation for missing values through semiparametric models. *Biometrics*, *67*, 799–809.

Copas, J. B., & Li, H. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society: Series B*, *59*, 55–95.

Covinsky, K. E., Palmer, R. M., Fortinsky, R. H., Counsell, S. R., Stewart, A. L., Kresevic, D., Burant, C. J., & Landefeld, C. S. (2003). Loss of independence in activities of daily living in older adults hospitalized with medical illnesses: Increased vulnerability with age. *Journal of the American Geriatrics Society*, *51*(4), 451–458.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., & Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, *52*, 111–125.

Curran, D., Molenberghs, G., Thijs, H., & Verbeke, G. (2004). Sensitivity analysis for pattern mixture models. *Journal of Biopharmaceutical Statistics*, *14*, 125–143.

Diodato, M., & Chedrawy, E. G. (2014). Coronary artery bypass graft surgery: The past, present, and future of myocardial revascularisation. *Surgery research and practice*, *2014*.

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, *25*, 88.

Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

Galimard, J.-E., Chevret, S., Curis, E., & Resche-Rigon, M. (2018). Heckman imputation models for binary or continuous mnar outcomes and mar predictors. *BMC Medical Research Methodology*, *18*, 1–13.

Gao, W., Hedeker, D., Mermelstein, R., & Xie, H. (2016). A scalable approach to measuring the impact of nonignorable nonresponse with an EMA application. *Statistics in Medicine*, *35*, 5579–5602.

Glonek, G. F. V. (1999). On identifiability in models for incomplete binary data. *Statistics & probability letters*, *41*, 191–197.

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters*, *99*(3), 585–590.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 153–161.

Heitjan, D. F., & Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics*, 2244–2253.

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195.

Jing, B., Qian, Y., Heitjan, D. F., & Xie, H. (2023). Tutorial: Assessing the impact of nonignorable missingness on regression analysis using index of local sensitivity to nonignorability. *Psychological Methods*.

Katz, S. (1963). The index of adl: A standardized measure of biological and psychosocial function. *J Am Med Assoc*, *185*, 914–919.

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine*, *17*, 2723–2732.

Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 209–225.

Lawrence, V. A., Hazuda, H. P., Cornell, J. E., Pederson, T., Bradshaw, P. T., Mulrow, C. D., & Page, C. P. (2004). Functional independence after major abdominal surgery in the elderly. *Journal of the American College of Surgeons*, *199*(5), 762–772.

Lee, T., & Shi, D. (2021). A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data. *Psychological Methods*, *26*, 466–485.

Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., & Carpenter, J. R. (2018). Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: A tutorial. *Pharmacoeconomics*, *36*, 889–901.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*, 125–134.

Little, R. J. A., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., et al. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, *367*, 1355–1360.

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Liu, T., & Heitjan, D. F. (2012). Sensitivity of the discrete-time kaplan–meier estimate to nonignorable censoring: Application in a clinical trial. *Statistics in Medicine*, *31*, 2998–3010.

Lord, D., Park, B.-J., & Model, P.-G. (2012). Negative binomial regression models and estimation methods. *Probability Density and Likelihood Functions. Texas A&M University, Korea Transport Institute*, 1–15.

Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Psychological Methods*, *25*, 157.

Ma, G., Troxel, A. B., & Heitjan, D. F. (2005). An index of local sensitivity to nonignorable drop-out in longitudinal modelling. *Statistics in Medicine*, *24*, 2129–2150.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2014). *Handbook of Missing Data Methodology*. CRC Press.

Mori, M., Djulbegovic, M., Hajduk, A. M., Holland, M. L., Krumholz, H. M., & Chaudhry, S. I. (2021). Changes in functional status and health-related quality of life in older adults after surgical, interventional, or medical management of acute myocardial infarction. *Seminars in thoracic and cardiovascular surgery*, *33*(1), 72–81.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal of applied econometrics*, *12*(3), 337–350.

Murphy, J. K., Xie, H., Nguyen, V. C., Chau, L. W., Oanh, P. T., Nhu, T. K., O'Neil, J., Goldsmith, C. H., Van Hoi, N., et al. (2020). Is supported self-management for depression effective for adults in community-based settings in Vietnam? A modified stepped-wedge cluster randomized controlled trial. *International Journal of Mental Health Systems*, *14*, 1–17.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *135*(3), 370–384.

Peng, C.-Y. J., Harwell, M., Liou, S.-M., Ehman, L. H., et al. (2006). Advances in missing data methods and implications for educational research. *Real Data Analysis*, *3178*.

Qian, Y. (2007). Do national patent laws stimulate domestic innovation in a global patenting environment? A cross-country analysis of pharmaceutical patent protection, 1978–2002. *The Review of Economics and Statistics*, *89*, 436–453.

Qian, Y., & Xie, H. (2011). No customer left behind: A distribution-free Bayesian approach to accounting for missing Xs in marketing models. *Marketing Science*, *30*, 717–736.

Qian, Y., & Xie, H. (2014). Which brand purchasers are lost to counterfeiters? An application of new data fusion approaches. *Marketing Science*, *33*, 437–448.

Qian, Y., & Xie, H. (2015). Driving more effective data-driven innovations: Enhancing the utility of secure databases. *Management Science*, *61*, 520–541.

Razzaq, F. A., Reyes, A. C., Tang, Q., Guo, Y., Liu, Y., et al. (2022). Life-long effects of malnutrition using semi-quantitative EEG analysis. *medRxiv*. https://doi.org/10.1101/2022.01.18.22269447

Rizopoulos, D., Verbeke, G., & Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, *95*, 63–74.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*, 1.

Schneider, S., Junghaenel, D. U., Ono, M., Broderick, J. E., & Stone, A. A. (2021). Iii. detecting treatment effects in clinical trials with different indices of pain intensity derived from ecological momentary assessment. *The Journal of Pain*, *22*, 386–399.

Serruys, P. W., Morice, M.-C., Kappetein, A. P., Colombo, A., Holmes, D. R., Mack, M. J., Ståhle, E., Feldman, T. E., Van Den Brand, M., Bass, E. J., et al. (2009). Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *New England journal of medicine*, *360*(10), 961–972.

Shin, T., Davison, M. L., & Long, J. D. (2017). Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality. *Psychological Methods*, *22*, 426.

Stineman, M. G., Xie, D., Pan, Q., Kurichi, J. E., Saliba, D., Rose, S. M. S.-F., & Streim, J. E. (2016). Understanding non-performance reports for instrumental activity of daily living items in population analyses: A cross-sectional study. *BMC geriatrics*, *16*, 1–7.

Troxel, A. B., Ma, G., & Heitjan, D. F. (2004). An index of local sensitivity to nonignorability. *Statistica Sinica*, 1221–1237.

Tsiatis, A. A. (2006). Semiparametric theory and missing data.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., & Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics*, *57*, 7–14.

Weiss, T., Carayannis, T., Jolly, R., Weiss, T., Ca, T., & Jolly, R. (2016). Missing data: A systematic review of how they are reported and handled. *Epidemiology*, *12*(5), 729–732.

Whitlock, E. L., Diaz-Ramirez, L. G., Smith, A. K., Boscardin, W. J., Covinsky, K. E., Avidan, M. S., & Glymour, M. M. (2021). Association of coronary artery bypass grafting vs percutaneous coronary intervention with memory decline in older adults undergoing coronary revascularization. *Jama*, *325*(19), 1955–1964.

Woodward, S. H., Jamison, A. L., Gala, S., Lawlor, C., Villasenor, D., Tamayo, G., & Puckett, M. (2021). Tracking positive and negative affect in PTSD inpatients during a service dog intervention. *Journal of Consulting and Clinical Psychology*, *89*, 551.

Xie, H. (2008). A local sensitivity analysis approach to longitudinal non-gaussian data with nonignorable dropout. *Statistics in Medicine*, *27*, 3155–3177.

Xie, H. (2009). Bayesian inference from incomplete longitudinal data: A simple method to quantify sensitivity to nonignorable dropout. *Statistics in Medicine*, *28*, 2725–2747.

Xie, H. (2012). Analyzing longitudinal clinical trial data with nonignorable missingness and unknown missingness reasons. *Computational Statistics & Data Analysis*, *56*, 1287–1300.

Xie, H., Gao, W., Xing, B., Heitjan, D. F., Hedeker, D., & Yuan, C. (2018). Measuring the impact of nonignorable missingness using the `R` package `isni`. *Computer Methods and Programs in Biomedicine*, 207–220.

Xie, H., & Heitjan, D. F. (2004). Sensitivity analysis of causal inference in a clinical trial subject to crossover. *Clinical Trials*, *1*, 21–30.

Xie, H., & Qian, Y. (2012). Measuring the impact of nonignorability in panel data with nonmonotone nonresponse. *Journal of Applied Econometrics*, *27*, 129–159.

Xie, H., Qian, Y., & Qu, L. (2011). A semiparametric approach for analyzing nonignorable missing data. *Statistica Sinica*, *21*, 1881–1899.

Yuan, C., Hedeker, D., Mermelstein, R., & Xie, H. (2020). A tractable method to account for high-dimensional nonignorable missing data in intensive longitudinal data. *Statistics in Medicine*, *39*, 2589–2605.

Zaninotto, P., & Falaschetti, E. (2011). Comparison of methods for modelling a count outcome with excess zeros: Application to activities of daily living (adl-s). *Journal of Epidemiology & Community Health*, *65*(3), 205–210.

Zhang, J., & Heitjan, D. F. (2006). A simple local sensitivity analysis tool for nonignorable coarsening: Application to dependent censoring. *Biometrics*, *62*, 1260–1268.

Zhang, J., & Heitjan, D. F. (2007). Impact of nonignorable coarsening on bayesian inference. *Biostatistics*, *8*, 722–743.

Zhang, Q., & Wang, L. (2017). Moderation analysis with missing data in the predictors. *Psychological Methods*, *22*, 649.

Appendix A. R codes for isniglm.nb()

```
isniglm.nb = function(formula, data, weights,  subset,  start=NULL) {

  ## (1) process the call and set up model frame.
  cl <- match.call()
  if (missing(data))  data <- environment(formula)

  mf <- match.call(expand.dots = FALSE)
  m <- match(c("formula", "data", "weights", "subset"), names(mf), 0L)
  mf <- mf[c(1L, m)]
  f <- Formula(formula)
  if (any(length(f)>2)) stop("Cannot have more than two models")
  mf$formula <- f
  options(na.action=na.pass)
  mf[[1L]] <- as.name("get_all_vars")  ## get original input variable
  mfvar<- eval(mf, parent.frame())
  if (!missing(subset)) {
    SubSet <-mfvar$subset
    mfvar <-mfvar[SubSet,]
  }
  mf[[1L]] <- as.name("model.frame")
  mf <- eval(mf, parent.frame())
  if (missing(weights)) WTs <- NULL else
    WTs <- model.extract(mf, "weights")
  if (!is.null(WTs) && !is.numeric(WTs))
    stop("'weights' must be a numeric vector")
  if (!is.null(WTs) && any(WTs < 0))
    stop("negative weights not allowed")
  if (is.null(WTs)){
    IWT <- 0
    WTs <- rep(1, nrow(mf))
  }
  isni_WTs <- NULL


  ##(2) Extract responses and predictors from the model frame
```

```
ymodel=formula(f, lhs=1, rhs=1)
x<- as.matrix(model.matrix(f, data=mf, rhs=1))
y<- model.part(f,data=mf, lhs=1)[,1]  ## will also be a vector
if (nrow(x)<1)  print("No predictor variables specified for the outcome")

s<- as.matrix(model.matrix(f, data=mf, rhs=length(f)[2]))
mf= cbind(mf, mfvar[!(names(mfvar) %in% names(mf))],  isni_WTs=WTs)

## check if the missing status indicator g is specified, if not add g in the lhs of formula
## and to the model.frame
if (length(f)[1]==1) {
  f <- update(f, .| g_ ~.)
  mf <- cbind(mf, g_=as.numeric(is.na(y)))
}
gmodel=formula(f, lhs=2,rhs=length(f)[2])
g<- mf$g_ <-  model.part(f,data=mf, lhs=2)[,1]
options(na.action=na.omit)

## Drop all observations with missing values in any outcome predictors  in X.
missX<- apply(x, 2, FUN=function (u) any(is.na(u)))
if (any(missX)) cat(paste("\n Note: There are missing values in fixed-effect
                          outcome predictors-- ",
  paste(dimnames(x)[[2]][missX], collapse=" "), ". Observations with missing values
  in these predictors will be dropped out from computing ISNI,
  which may reduce ISNI values if the outcome is missing concurrently.
  Alternatively one can impute these missing predictors values and re-compute ISNI. ",
  sep=""))
missX  <- apply(x, 1, FUN=function (u) any(is.na(u)))
WTs=WTs[!missX]
#OffSets = OffSets[!missX]
s=s[!missX, ,drop=F]
g=g[!missX]
mf=mf[!missX,,drop=F]
options(na.action=na.pass)
x<- as.matrix(model.matrix(f, data=mf, rhs=1))
y<- model.part(f,data=mf, lhs=1)[,1]  ## will also be a vector
xomit <-  as.matrix(model.matrix(f, data=mf[!is.na(y), ], rhs=1))
if (ncol(x) != ncol(xomit)) {
  cat(paste("\n All variable names in the design matrix of ISNI analysis including
      observations with missing outcomes: \n", paste (dimnames(x)[[2]], collapse=" ")))
  cat(paste("\n All variable names in the design matrix of the MAR model excluding
      observations with missing outcomes:  \n", paste (dimnames(xomit)[[2]], collapse=" ")))
  stop("\n The design matrix for the MAR model and ISNI analysis are different as
      shown above. Please modify your model formula specification,
      e.g. avoiding using as.factor function in model formula.  ")
```

```
}
options(na.action=na.omit)

yo<-y[g==0]
xo<-as.matrix(x[g==0,,drop=F])
xm<-as.matrix(x[g==1,,drop=F])
sdy<-sd(as.numeric(yo))

## fitting a glm outcome model on fully observed data
rego <- glm.nb(ymodel, data=mf,weights=isni_WTs,start=start, x=TRUE)
coef<-summary(rego)$coef[,1]
yfit<-rego$fitted.values

nabla11<-obs.cov.matrix(rego)
print(nabla11)


# calculating the second part of ISNI

## Fit missing data model, drop any predictors with missing values from the model.
missS  <- apply(s, 2, FUN=function (u) any(is.na(u)))
s <- as.data.frame(cbind(s, isni_WTs=mf$isni_WTs, g_=mf$g_))
regzg<-multinom(g_ ~.-1-isni_WTs, data=s[,!missS],   weights=isni_WTs)
hm<-1-regzg$fitted.values[g==1]
cc=as.numeric(exp(x%*%coef))
ccm<-WTs[g==1]*hm*cc[g==1]
sscp2<-t(xm)%*%ccm
alpha<-0
nabla13<-c(sscp2,alpha)
print(nabla13)



# calculating ISNI
isni<-nabla11%*%nabla13
# calculate MinNI
se.beta<-summary(rego)$coefficient[,2]
se.alpha<-rego$SE.theta
se<-c(se.beta,theta=se.alpha)

## Per Prof. Xie, adding the standard deviation to the MinNI calcualtion

SD_y<-sd(yo)

MinNI<-abs((SD_y*se)/isni)
```

```
  ##Make the output list
  coeff<-c(rego$coef,theta=rego$theta)
  pvalue<-c(summary(rego)$coefficient[,4],NULL)
  res=list(coef=coeff,se=se,pvalue=pvalue,isni=isni, MinNI=MinNI, call=cl,
           formula=formula, deviance=rego$deviance, aic=rego$aic,
           nabla11=nabla11, nabla13=nabla13)
  class(res) = c(res$class, "isniglm.nb")
  res
}

## Getting the observed information matrix -- calculate the full variance covariance

obs.cov.matrix <- function(rego) {

  #p is number of regression coefficients
  p <- dim(vcov(rego))[1]

  #construct observed covariance matrix
  covmat <- array(0, dim=c(p+1, p+1))

  #1.calculate the second derivatives among betas
  for (i in 1:p) {
    for (j in 1:p) {
      covmat[i,j] <- sum((1+rego$y/rego$theta)*rego$fitted.values*rego$x[,i]*rego$x[,j]
            / (1+rego$fitted.values/rego$theta)^2)
    }
  }

  #2. calculate the second derivatives for beta and alpha
  for (i in 1:p) {
    covmat[(p+1),i] <- -sum(((rego$y-rego$fitted.values) * rego$fitted.values
                / ( (rego$theta+rego$fitted.values)^2 )) * rego$x[,i] )
    covmat[i,(p+1)] <- covmat[(p+1),i]
  }
  #3. Calculate the second derivative for alphas
  covmat[(p+1),(p+1)] <- -sum(trigamma(rego$theta+rego$y) -
     trigamma(rego$theta) -
     1/(rego$fitted.values+rego$theta) +
     (rego$theta+rego$y)/(rego$theta+rego$fitted.values)^2 -
     1/(rego$fitted.values+rego$theta) +
     1/rego$theta)
  #return variance-covariance matrix
  solve(covmat)
}
```

```
summary.isniglm.nb<-function(object, digits = max(3, getOption("digits") - 2),
                          ...) {

  if (class(object) != "isniglm.nb")  stop('Invalid object class')
  cat("\nCall:\n", paste(deparse(object$call), sep = "\n",
                        collapse = "\n"), "\n\n", sep = "")
  ## Name the columns
  isniname<-c('MAR Est.','Std. Err','Pr(>|z|)','ISNI','MinNi')

  ## Set up matrix to hold result
  res<-matrix(0,length(object$coef),length(isniname))
  dimnames(res)<-list(names(object$coef),isniname)

  for (i in 1:length(object$coef))
    res[i,]<- c(object$coef[i],object$se[i],object$pvalue[i],object$isni[i],object$MinNI[i])
  printCoefmat(res, digits = digits,  cs.ind = 1:2)

}

print.isniglm.nb<-function(x, digits = max(3, getOption("digits") - 2), ...) {

  cat("\nCall:\n", paste(deparse(x$call), sep = "\n",
                        collapse = "\n"), "\n\n", sep = "")
  if (class(x) != "isniglm.nb")  stop('Invalid object class')

  if (length(x$coef)>0) {
    cat("ISNIs:\n")
    print.default(format(x$isni, digits = digits), print.gap = 2L,
                  quote = FALSE)
    cat("\n")
    cat("c statistics:\n")
    print.default(format(x$c, digits = digits), print.gap = 2L,
                  quote = FALSE)
  }
  else cat("No coefficients\n")
  cat("\nResidual Deviance of the MAR model: ",
    paste(format(x$deviance, digits=digits),
    sep = "\n", collapse = "\n"), "\n", sep = "")
  cat("\nAIC of the MAR model: ", paste(format(x$aic, digits=digits), sep = "\n",
                                        collapse = "\n"), "\n", sep = "")
  cat("\n")
  invisible(x)

}
```

Appendix B. R codes for simulation study 2 on various of dispersion

```
set.seed(314159)
n_datasets <- 1000  # Number of datasets per size

# Initialize lists to store datasets
datasets <- vector("list", Upper * n_datasets)

# Data generation loop
for (size in 1:Upper) {
  for (j in 1:n_datasets) {
    n <- 500
    x1 <- rnorm(n, mean = 0, sd = 1)
    eta <- 1 + 0.5 * x1
    mu <- exp(eta)
    y <- rnbinom(n, size = size, mu = mu)
    missing_prob <- plogis(-2 + 2 * x1 + 0.1 * y)
    G <- rbinom(n, 1, missing_prob)
    yy <- y
    yy[G == 1] <- NA
    datasets[[(size - 1) * n_datasets + j]] <- data.frame(y, yy, x1)
  }
}

# Initialize vectors to store results
ISNI_nb_intercept <- ISNI_pos_intercept <- ISNI_nb_x1 <- ISNI_pos_x1 <- vector("list", Upper)
MinNI_nb_intercept <- MinNI_pos_intercept <- MinNI_nb_x1 <-
                        MinNI_pos_x1 <- vector("list", Upper)
MAR_nb_intercept <- MAR_pos_intercept <- MAR_nb_x1 <- MAR_pos_x1 <- vector("list", Upper)

ymodel <- yy ~ x1

for (size in 1:Upper) {
  nb_intercept <- pos_intercept <- nb_x1 <- pos_x1 <- numeric(n_datasets)
  minni_nb_intercept <- minni_pos_intercept <- minni_nb_x1 <-
```

```
                                                minni_pos_x1 <- numeric(n_datasets)
  mar_nb_intercept <- mar_pos_intercept <- mar_nb_x1 <- mar_pos_x1 <- numeric(n_datasets)

  for (j in 1:n_datasets) {
    data <- datasets[[(size - 1) * n_datasets + j]]

    # Fit Negative Binomial model
    fit.nb <- isniglm.nb(ymodel, data = data)
    nb_intercept[j] <- summary(fit.nb)[1,4]
    nb_x1[j] <- summary(fit.nb)[2,4]
    minni_nb_intercept[j] <- summary(fit.nb)[1,5]
    minni_nb_x1[j] <- summary(fit.nb)[2,5]

    mar_nb_intercept<- summary(fit.nb)[1,1]
    mar_nb_x1<- summary(fit.nb)[2,1]

    # Fit Poisson model
    fit.pos <- isniglm.test(ymodel, data = data, family = "poisson")
    pos_intercept[j] <- summary(fit.pos)[1,3]
    pos_x1[j] <- summary(fit.pos)[2,3]
    minni_pos_intercept[j] <- summary(fit.pos)[1,4]
    minni_pos_x1[j] <- summary(fit.pos)[2,4]

    mar_pos_intercept[j] <- summary(fit.pos)[1,1]
    mar_pos_x1[j] <- summary(fit.pos)[2,1]

  }

  # Store the aggregated results
  ISNI_nb_intercept[[size]] <- nb_intercept
  ISNI_pos_intercept[[size]] <- pos_intercept
  ISNI_nb_x1[[size]] <- nb_x1
  ISNI_pos_x1[[size]] <- pos_x1

  MinNI_nb_intercept[[size]] <- minni_nb_intercept
  MinNI_pos_intercept[[size]] <- minni_pos_intercept
  MinNI_nb_x1[[size]] <- minni_nb_x1
  MinNI_pos_x1[[size]] <- minni_pos_x1

  MAR_nb_intercept[[size]] <- mar_nb_intercept
  MAR_pos_intercept[[size]] <- mar_pos_intercept
  MAR_nb_x1[[size]] <- mar_nb_intercept
  MAR_pos_x1[[size]] <- mar_pos_x1
}
```

```
calculate_ci <- function(data) {
  mean_data <- mean(data)
  sd_data <- sd(data)
  ci_lower <- mean_data - 1.96 * sd_data
  ci_upper <- mean_data + 1.96 * sd_data
  return(c(mean = mean_data, lower = ci_lower, upper = ci_upper))
}

# Initialize data frames to store the aggregated results
ISNI_nb_int<- ISNI_pos_int <- MinNI_nb_int <- MinNI_pos_int <-
 data.frame(size = integer(), mean = numeric(), lower = numeric(), upper = numeric())
ISNI_nb_x1a <- ISNI_pos_x1a <- MinNI_nb_x1a <- MinNI_pos_x1a <-
 data.frame(size = integer(), mean = numeric(), lower = numeric(), upper = numeric())
nb_int <- pos_int <- nb_x1a <- pos_x1a <-
 data.frame(size = integer(), mean = numeric(), lower = numeric(), upper = numeric())

# Loop through each size
for (size in 1:Upper) {

  # Calculate mean and CI for each parameter and model for intercepts
  ISNI_nb_int <- rbind(ISNI_nb_int, c(size = size,
                    calculate_ci(unlist(ISNI_nb_intercept[[size]]))))
  ISNI_pos_int <- rbind(ISNI_pos_int, c(size = size,
                    calculate_ci(unlist(ISNI_pos_intercept[[size]]))))
  MinNI_nb_int <- rbind(MinNI_nb_int, c(size = size,
                    calculate_ci(unlist(MinNI_nb_intercept[[size]]))))
  MinNI_pos_int <- rbind(MinNI_pos_int, c(size = size,
                calculate_ci(unlist(MinNI_pos_intercept[[size]]))))

  # Calculate mean and CI for each parameter and model for x1 variable
  ISNI_nb_x1a <- rbind(ISNI_nb_x1a, c(size = size,
                    calculate_ci(unlist(ISNI_nb_x1[[size]]))))
  ISNI_pos_x1a <- rbind(ISNI_pos_x1a, c(size = size,
                    calculate_ci(unlist(ISNI_pos_x1[[size]]))))
  MinNI_nb_x1a <- rbind(MinNI_nb_x1a, c(size = size,
                    calculate_ci(unlist(MinNI_nb_x1[[size]]))))
  MinNI_pos_x1a <- rbind(MinNI_pos_x1a, c(size = size,
                    calculate_ci(unlist(MinNI_pos_x1[[size]]))))

  # Calculate mean estimates and CI intercept and X1
  nb_int <- rbind(nb_int, c(size = size,
                    calculate_ci(unlist(MAR_nb_intercept[[size]]))))
  pos_int <- rbind(pos_int, c(size = size,
                    calculate_ci(unlist(MAR_pos_intercept[[size]]))))
  nb_x1a <- rbind(nb_x1a, c(size = size,
```

```
                    calculate_ci(unlist(MAR_nb_x1[[size]])))
  pos_x1a <- rbind(pos_x1a, c(size = size,
                    calculate_ci(unlist(MAR_pos_x1[[size]]))))


}

# Naming columns for clarity
cols <- c("size", "mean", "lower", "upper")
colnames(ISNI_nb_int) <- colnames(ISNI_pos_int) <- colnames(MinNI_nb_int) <-
colnames(MinNI_pos_int) <- cols
colnames(ISNI_nb_x1a) <- colnames(ISNI_pos_x1a) <- colnames(MinNI_nb_x1a) <-
colnames(MinNI_pos_x1a) <- cols
colnames(nb_int) <- colnames(pos_int) <- colnames(nb_x1a) <- colnames(pos_x1a) <- cols

# Prepare data for plotting
ISNI_nb_int$Model <- "Negative Binomial"
ISNI_pos_int$Model <- "Poisson"
ISNI_nb_x1a$Model <- "Negative Binomial"
ISNI_pos_x1a$Model <- "Poisson"
MinNI_nb_int$Model <- "Negative Binomial"
MinNI_pos_int$Model <- "Poisson"
MinNI_nb_x1a$Model <- "Negative Binomial"
MinNI_pos_x1a$Model <- "Poisson"

# Combine data for both models
ISNI_combined_int<-rbind(ISNI_nb_int,ISNI_pos_int)
ISNI_combined_x1 <- rbind(ISNI_nb_x1a, ISNI_pos_x1a)
MinNI_combined_int <- rbind(MinNI_nb_int, MinNI_pos_int)
MinNI_combined_x1 <- rbind(MinNI_nb_x1a, MinNI_pos_x1a)

# Plotting
plot_ISNI_int <- ggplot(ISNI_combined_int, aes(x = size, y = mean, color = Model)) +
  geom_line() +
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = Model), alpha = 0.2) +
  labs(title = "ISNI for Intercept",
       x = "Alpha", y = "Intercept ISNI values") +
  theme_minimal()
plot_ISNI_x1 <- ggplot(ISNI_combined_x1, aes(x = size, y = mean, color = Model)) +
  geom_line() +
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = Model), alpha = 0.2) +
  labs(title = "ISNI for X1",
       x = "Alpha", y = "X1 ISNI values") +
  theme_minimal()
plot_MinNI_int <- ggplot(MinNI_combined_int, aes(x = size, y = mean, color = Model)) +
```

```
  geom_line() +
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = Model), alpha = 0.2) +
  labs(title = "MinNI for Intercept",
       x = "Alpha", y = "Intercept MinNI Values") +
  theme_minimal()
plot_MinNI_x1 <- ggplot(MinNI_combined_x1, aes(x = size, y = mean, color = Model)) +
  geom_line() +
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = Model), alpha = 0.2) +
  labs(title = "MinNI for X1",
       x = "Alpha", y = "X1 MinNI Value") +
  theme_minimal()
# Combining all plots
plots.isni.minni <- list(plot_ISNI_int, plot_ISNI_x1,plot_MinNI_int,plot_MinNI_x1 )

combined_plot1 <- ggarrange(plotlist = plots.isni.minni,
                                common.legend = TRUE, legend = "top")
print(combined_plot1)
```

Appendix C. R codes for simulation study 3 on various degrees of missing proportion

```
## Function to find the intercept from the missing data model that
## gives a specific mean

find_intercept <- function(intercept, target_mean) {
  missing_prob <- plogis(intercept + 2 * x1 + 0.5*y)
  G <- rbinom(n, 1, missing_prob)
  return(mean(G) - target_mean)
}

# Generate data
set.seed(314159)
n <- 500
x1 <- rnorm(n, mean = 0, sd = 1)
eta <- 1 + 0.5 * x1
mu <- exp(eta)
size <- 2
y <- rnbinom(n, size = size, mu = mu)

# Initialize vectors to store intercepts and means
intercepts <- numeric()
means <- numeric()

# Loop through target means
for(target_mean in seq(0.1, 0.8, by = 0.1)){
  # Find the root of the function
  intercept <- uniroot(find_intercept, lower = -10, upper = 10,
                       target_mean = target_mean)$root
  # Store the intercept and the corresponding mean
  intercepts <- c(intercepts, intercept)
  means <- c(means, target_mean)
}
```

```
# Create a data frame to store the results
df_results <- data.frame(target_mean = means, intercept = intercepts)

data_list <- list()

data_counter <- 1  # Initialize a counter to keep track of total data frames

# Loop through target means
for(i in 1:length(intercepts)) {
  # Get the intercept from the intercepts vector
  intercept <- intercepts[i]

  for(j in 1:1000){
    # Generate missing probabilities and binary indicator for each dataset
    missing_prob <- plogis(intercept + 2 * x1 + 0.5*y)
    G <- rbinom(n, 1, missing_prob)

    # Replace missing values
    yy <- y
    yy[G==1] <- NA

    # Generate data frame and add to list
    # Include target mean in the data frame
    data <- data.frame(y = y, yy = yy, x1 = x1, miss_porportion = target_mean[i])
    data_list[[data_counter]] <- data
    data_counter <- data_counter + 1  # Increment the counter
  }
}

# Initialize lists to store models
model_summaries.nb.miss <- list()
model_summaries.pos.miss <- list()

# Initialize lists to store coefficients
ISNI_nb.intercept.miss <- list()
ISNI_pos.intercept.miss <- list()
ISNI_nb.x1.miss <- list()
ISNI_pos.x1.miss <- list()

ymodel <- as.formula("yy ~ x1")

# Fit models and get summaries
for (i in 1:length(data_list)){
  data <- data_list[[i]]
```

```
  fit.nb <- isniglm.nb(ymodel, data=data)
  fit.pos <- isniglm.test(ymodel, data=data, family="poisson")
  model_summaries.nb.miss[[i]] <- summary(fit.nb)
  model_summaries.pos.miss[[i]] <- summary(fit.pos)
}

# Extract the desired coefficients
for (i in 1:length(data_list)) {
  ISNI_nb.intercept.miss[[i]] <- model_summaries.nb.miss[[i]][1, 4]
  ISNI_pos.intercept.miss[[i]] <- model_summaries.pos.miss[[i]][1, 3]
  ISNI_nb.x1.miss[[i]] <- model_summaries.nb.miss[[i]][2,4]
  ISNI_pos.x1.miss[[i]] <-model_summaries.pos.miss[[i]][2,3]
}

# Convert lists to data frames
ISNI_nb_df.intercept.miss <-
    data.frame(miss_proportion = rep(seq(0.1, 0.8, by = 0.1),each = 1000),
                Intercept = unlist(ISNI_nb.intercept.miss),
                model=rep("NB",length(data_list)))
ISNI_pos_df.intercept.miss <-
    data.frame(miss_proportion = rep(seq(0.1, 0.8, by = 0.1), each = 1000),
                Intercept = unlist(ISNI_pos.intercept.miss),
                model=rep("Poisson",length(data_list)))
ISNI_nb_df.x1.miss <-
    data.frame(miss_proportion = rep(seq(0.1, 0.8, by = 0.1), each = 1000),
               X1 = unlist(ISNI_nb.x1.miss),
               model=rep("NB",length(data_list)))
ISNI_pos_df.x1.miss <-
    data.frame(miss_proportion = rep(seq(0.1, 0.8, by = 0.1), each = 1000),
               X1 = unlist(ISNI_pos.x1.miss),
               model=rep("Poisson",length(data_list)))

# Combine data frames
ISNI_intercept.miss <- rbind(ISNI_nb_df.intercept.miss, ISNI_pos_df.intercept.miss)
ISNI_X1.miss <- rbind(ISNI_nb_df.x1.miss, ISNI_pos_df.x1.miss)

#### now let's plotting the confidence band;
ISNI_intercept.miss1 <- ISNI_intercept.miss %>%
  group_by(miss_proportion, model) %>%
  summarise(Intercept_mean = mean(Intercept),
            Intercept_se = sd(Intercept) / sqrt(n()))

ISNI_X1.miss1 <- ISNI_X1.miss %>%
  group_by(miss_proportion, model) %>%
  summarise(X1_mean = mean(X1),
```

```
            X1_se = sd(X1) / sqrt(n()))

# Now, plot the mean and add the confidence interval using geom_ribbon
# a. intercept
plot.int.miss <- ggplot(ISNI_intercept.miss1,
  aes(x = miss_proportion, y = Intercept_mean, color = model)) +
  geom_line(aes(y = Intercept_mean), linetype = "dashed") +
  geom_point(aes(y = Intercept_mean), shape = 17, size = 2) +
  geom_ribbon(aes(ymin = Intercept_mean - 1.96*Intercept_se,
                  ymax = Intercept_mean + 1.96*Intercept_se),
              alpha = 0.2) +
  labs(title = NULL,
       x = "Missing Proportion",
       y = "Intercept") +
  scale_color_discrete(labels = c("Negative Binomial", "Poisson"))
print(plot.int.miss)

# b. X1
plot.X1.miss <- ggplot(ISNI_X1.miss1,
  aes(x = miss_proportion, y = X1_mean, color = model)) +
  geom_line(aes(y = X1_mean), linetype = "dashed") +
  geom_point(aes(y = X1_mean), shape = 18, size = 2) +
  geom_ribbon(aes(ymin = X1_mean - 1.96*X1_se,
                  ymax = X1_mean + 1.96*X1_se),
              alpha = 0.2) +
  labs(title = NULL,
       x = "Missing Proportion",
       y = "X1") +
  scale_color_discrete(labels = c("Negative Binomial", "Poisson"))
print(plot.X1.miss)

## combining the missing plot;
plots.miss <- list(plot.int.miss, plot.X1.miss)

combined_plot.miss <- ggarrange(plotlist = plots.miss,
                                common.legend = TRUE, legend = "top")
print(combined_plot.miss)
```