Summary Measures for Quantifying the Extent of Visit Irregularity in Longitudinal Data

by

Armend Lokku

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Biostatistics
University of Toronto

© Copyright 2020 by Armend Lokku

Abstract

Summary Measures for Quantifying the Extent of Visit Irregularity in Longitudinal Data

Armend Lokku
Doctor of Philosophy
Graduate Department of Biostatistics
University of Toronto
2020

A common feature of observational longitudinal data is variability in the timings of visits across individuals. Irregular visits merit careful scrutiny as they can be associated with the outcome trajectory and lead to biased results if not properly addressed. Similar to missing data, exploring the extent of irregularity is important because it can help determine whether specialized methods for irregular data are needed. However, there are currently no measures for quantifying the extent of irregularity. This thesis proposes visual and numerical measures of irregularity which are based on bins constructed across the study duration. To visually assess the extent of irregularity, the bin widths are varied and the mean proportions of individuals with 0, 1, and >1 visits per bin are plotted. With perfect repeated measures, the mean proportions of individuals with 1 visit per bin are 1, and the mean proportions of individuals with 0 and >1 visits per bin are 0. Missingness leads to individuals with 0 visits per bin while irregularity leads to both individuals with >1 visit per bin and individuals with 0 visits per bin. To numerically assess the extent of irregularity, we use the area under the curve (AUC) of the mean proportions of individuals with 0 visits per bin plotted against the mean proportions of individuals with >1 visit per bin. To be an effective measure of irregularity, the AUC should increase with increasing irregularity, be intuitive, and be invariant to sample size and follow-up length. Ideally, the AUC should be invariant to missingness, but the AUC increases as the level of missingness increases due to increased proportions of individuals with 0 visits per bin. To allow judgement solely on irregularity and not missingness, likelihood-based estimation will be used to derive an AUC assuming no missingness. Simulation results demonstrate that the AUC increases with increasing irregularity and is invariant to sample size and the number of scheduled measurement occasions. Overall, these measures can lead to an improved quality of statistical analyses by informing the modelling approach. Health administrative data are increasingly available, but are susceptible to irregular visit timings. It is thus crucial to assess the extent of irregularity to inform the need for specialized methods for irregular data that minimize the potential for biased results.

Acknowledgements

I would like to thank my supervisor, Dr. Eleanor Pullenayegum, for mentoring me and providing me with the best possible PhD experience. She is an excellent motivator who challenged me intellectually, and I would not be the person I am today without her. She was always available for help and has a gift for making really complex problems intuitive and straightforward to understand. She is brilliant, personable and very understanding, and I was truly lucky to have her as a supervisor.

I would also like to thank my thesis committee members Dr. Wendy Lou and Dr. Olli Saarela. I will always be thankful to Dr. Wendy Lou for admitting me to the PhD program. It's amazing how dedicated she is to the students and how she always finds the time to sit with you and give advice and wonderful support, she is truly inspirational. Dr. Olli Saarela was one of the best professors I had during my graduate studies, and he gave me great suggestions for strengthening my thesis so I am indebted to him as well. I thoroughly enjoyed my time in the Biostatistics department and I met people who I feel honoured to call my friends and colleagues. I would like to specifically thank Mohsen Soltanifar and Shahriar Shams for their support and friendship during my degree.

To my family, I would like to thank them for their unwavering support and love. I am blessed to have the greatest and most caring mother in the world. Her support and love is unparalleled and this degree would not have been possible without her. My father deserves special credit for nurturing me to be academically inclined and challenging me to fulfill my potential. I would also like to thank my brother Fisnik, my sister Rona, and the nicest person in the world Mirjeta for their love and support, I am truly grateful to have them in my life.

Lastly, I would like to thank the love of my life and my best friend Sara. She deserves special praise for her profound love, patience and for being my rock. She made my graduate experience even better by filling my days with fun and laughter and accompanying me on our memorable adventures. She is the greatest person I've ever met and I can't wait to spend the rest of my life with her.

The TARGet Kids! research cohort provided data for the manuscripts in Chapters 2 and 3 as submitted. We thank all of the participating families for their time and involvement in TARGet Kids! and are grateful to all practitioners who are currently involved in the TARGet Kids! practice-based research network.

The Sequenced Treatment Alternatives to Relieve Depression study obtained from the National Institute of Mental Health provided data for Chapter 5. Data and/or research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health (NIMH) Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health. Dataset identifier(s): [NIMH Data Archive Collection ID(s) or NIMH Data Archive Digital Object Identifier (10.15154/1518466)]. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDA.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Longitudinal Data

Longitudinal data occur when measurements are collected on individuals repeatedly over time. In contrast, cross-sectional data provide a snapshot of variables at a single point in time. Longitudinal data can be used to assess changes within individuals over time while cross-sectional data cannot. To illustrate this, suppose the i^{th} individual ($i = 1, 2 \dots n$) has their outcome measured at times $t_{i1}, \dots, t_{iJ_i}$ such that each time t satisfies $0 \leq t \leq \tau$ (where τ is the total study duration), and let C_i denote their censoring time. For the i^{th} individual at time t, let $Y_i(t)$ denote the outcome, let $\epsilon_i(t)$ denote the random error where the vector $\boldsymbol{\epsilon}_i = (\epsilon_i(t_{i1}), \dots, \epsilon_i(t_{iJ_i}))$ has some multivariate distribution with mean 0 and covariance matrix R_i, and let $\boldsymbol{X}_i(t)$ denote a p-dimensional vector of covariates (it is assumed that covariates are always observed). Cross-sectional studies are restricted to models of the form:

$$Y_i(t_{i1}) = \boldsymbol{X}_i(t_{i1})'\boldsymbol{\beta_C} + \epsilon_i(t_{i1}) \tag{1.1}$$

where the p-dimensional vector of regression coefficients $\boldsymbol{\beta_C}$ represents the change in the mean of $Y_i(t_{i1})$ associated with a unit increase in $\boldsymbol{X}_i(t_{i1})$ [1]. With longitudinal studies, models can be extended to incorporate effects of changes in covariates over time within an individual:

$$Y_i(t) = \boldsymbol{X}_i(t_{i1})'\boldsymbol{\beta_C} + (\boldsymbol{X}_i(t) - \boldsymbol{X}_i(t_{i1}))'\boldsymbol{\beta_L} + \epsilon_i(t) \tag{1.2}$$

This model reduces to the cross-sectional case when $t = t_{i1}$ (1.1). Let the p-dimensional vector of regression coefficients $\boldsymbol{\beta_L}$ represents the change in the mean of $Y_i(t) - Y_i(t_{i1})$ associated with a unit increase in $\boldsymbol{X}_i(t) - \boldsymbol{X}_i(t_{i1})$:

$$E(Y_i(t) - Y_i(t_{i1})|\boldsymbol{X}_i(t), \boldsymbol{X}_i(t_{i1})) = (\boldsymbol{X}_i(t) - \boldsymbol{X}_i(t_{i1}))'\boldsymbol{\beta_L}$$

Cross-sectional studies can evaluate the association between changes in covariates over time and changes in the mean outcome within an individual, but it must be assumed that $\boldsymbol{\beta_L} = \boldsymbol{\beta_C}$ [1]. Longitudinal studies do not require this unlikely assumption as both parameters can be estimated separately. The Framingham heart study [2] is an example of an ongoing longitudinal cohort study which has identified

risk factors (e.g. genetics, diet, exercise etc.) for developing cardiovascular disease and hypertension later in life. Another example of an impactful longitudinal study is the Treatment of Lead-Exposed Children [3] randomized controlled trial which demonstrated that the use of Succimer lowered mean blood lead levels over time compared to placebo in children living in homes with lead-based paint.

Observational longitudinal data are readily available and are a pragmatic alternative to randomized controlled trials which can be costly and not always feasible. Furthermore, there is an increasing amount of high volume "Big Data" in health-care which are aggregated from multiple sources. For example, electronic health records are obtained from various organizational databases such as hospitals, laboratories, walk-in clinics etc. Big Data can be a beneficial tool for assessing population-level hypotheses; however, it can be difficult to analyze the data and maintain overall data quality due to the high volume of information which is frequently collected, and the way in which it is collected [4]. For example, the Ontario Drug Benefit Program (ODB) [5] is Ontario's funded drug program which contains information on opioid prescriptions across eligible individuals. However, prescription data can be misleading if used to evaluate actual opioid usage as it is not uncommon for individuals who misuse opioids to misrepresent their opioid usage to doctors [6]. Observational data generally present a set of challenges including confounding, missing data, and irregular observation times which can lead to bias if not properly addressed.

The topic of my thesis is irregular longitudinal data within an observational setting. For the purpose of this thesis, outcomes are assumed to be measured at times when individuals visit with a care-provider. Irregular data describes the scenario where the timings of visits vary across individuals. For example, Figure 1.1 displays the timings of visits for 20 randomly selected individuals from a study on childhood-onset Systemic Lupus Erythematosus (cSLE) [7] which recommended at least 1 visit every 6 months; however, it was suspected that individuals visited more frequently when their disease status worsened.

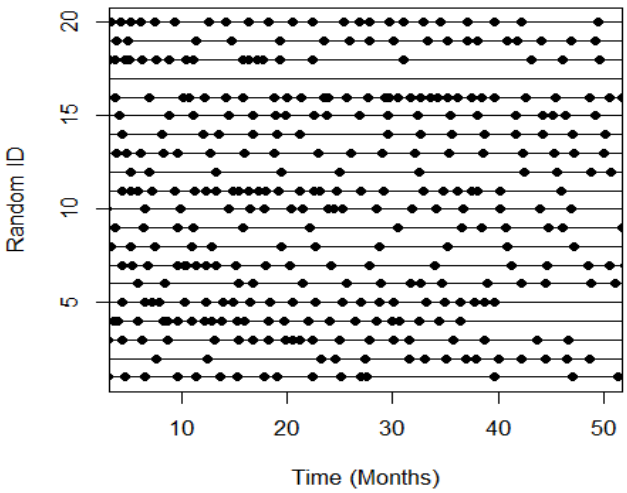

Figure 1.1: The visit timings for a random subset of 20 individuals from the cSLE study.

Irregular visit times are problematic because they can be associated with the outcome trajectory and lead to biased estimates if neglected. For example, newborn infants with slow weight gain will

be monitored by their doctors more frequently, and thus visit times with lower weights will be over-represented in the data, and the overall mean of an outcome relating to weight (e.g. body mass index) will be underestimated unless the visit process is addressed. Even when the study is intended to be repeated measures, individuals may miss scheduled visits, have delayed scheduled visits, and show up for unscheduled visits. An example of this is the TARGet Kids! cohort study [8] where although children were invited to visit a clinic at ages 2, 4, 6, 9, 12, 15, 18, 24 months and then annually thereafter, parents also brought their children for as-needed "sick" visits. It can be difficult to determine at what point deviations from protocol are too large to treat data as repeated measures. This is important because if it is deemed that there is a risk of bias by treating the data as repeated measures, then specialized methods for irregular data should be considered.

1.2 Methods

1.2.1 Standard Methods for Longitudinal Data

A common objective of longitudinal studies is to characterize changes in the study outcome over time. Longitudinal data are clustered because measurements are repeatedly collected on the same individual at different times. The dependence among an individual's measurements is accounted for using techniques for correlated data. Two main regression approaches are generalized estimating equations (GEEs) [9] and mixed effects models [10]. Carrying over the notation from Section 1.1, generalized linear models can be used to describe the marginal mean outcome through a chosen link function $g()$ ($\mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta}) = E(Y_i(t)|\boldsymbol{X}_i(t)) = g^{-1}(\boldsymbol{X}_i(t)'\boldsymbol{\beta})$) where $\boldsymbol{\beta}$ denotes a p-dimensional vector of regression coefficients.

The motivation for generalized estimating equations is based on generalized least squares (GLS) estimation of $\boldsymbol{\beta}$ with an identity link function where the following is minimized:

$$\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) \tag{1.3}$$

where $\boldsymbol{\mu}_i$ is a vector of $\mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})$, $\boldsymbol{y}_i$ is a vector of outcomes $Y_i(t)$, and Σ_i is the covariance matrix of the outcomes. If a minimum of this function exists, it must solve the following:

$$\sum_{i=1}^{n} \boldsymbol{X}_i(t)' \Sigma_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0 \tag{1.4}$$

With generalized estimating equations, a similar function must be solved to estimate $\boldsymbol{\beta}$:

$$\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0 \tag{1.5}$$

where $\boldsymbol{D}_i = \frac{d\boldsymbol{\mu}_i}{d\boldsymbol{\beta}}$, and $\boldsymbol{V}_i$ is the working covariance matrix which has the form:

$$\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2} Corr(\boldsymbol{y}_i) \boldsymbol{A}_i^{1/2} \tag{1.6}$$

where $\boldsymbol{A}_i$ is a diagonal matrix with $\phi v(g(E(Y_i(t)|\boldsymbol{X}_i(t))))$ along the diagonal ($v()$ represents a known variance function and ϕ denotes a non-negative scale parameter), and $Corr(\boldsymbol{y}_i)$ is the correlation matrix

of the outcomes in which a structure for pair-wise correlations is specified (e.g. exchangeable structure, independence structure etc.). Generalized estimating equations rely on the assumption that there is a linear relationship between the covariates and the transformed mean. To implement this method, it is necessary to specify the within-individual correlation structure, and the relationship between the first two moments. For example, if the study outcome is binary, then the first two moments would be specified in accordance with the Bernoulli distribution. The sandwich estimator [11] [12] is used to estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$:

$$\left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1} \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right) \left(\sum_{i=1}^{n} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1} \tag{1.7}$$

where $\hat{\boldsymbol{D}}_i$ and $\hat{\boldsymbol{\mu}}_i$ denote the respective $\boldsymbol{D}_i$ and $\boldsymbol{\mu}_i$ vectors evaluated at $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{V}}_i$ denotes the $\boldsymbol{V}_i$ matrix evaluated at $\hat{\boldsymbol{\beta}}$, the estimate of the scale parameter ϕ, and the parameter estimates of the specified correlation matrix of the outcomes $Corr(\boldsymbol{y}_i)$.

With mixed effects models, the mean outcome at the individual level is modelled as a combination of population level fixed effects and individual level random effects. Let $\boldsymbol{Z}_i(t)$ denote a q-dimensional vector of covariates for the random effects (it is assumed that covariates are always observed), and let $\boldsymbol{u}_i$ denote a q-dimensional vector of random effects. The random effects $\boldsymbol{u}_i$ have a multivariate normal distribution with mean 0 and covariance matrix G, and are assumed to be independent of the covariates. The linear mixed effects models can be expressed as:

$$Y_i(t) = \boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_i + \epsilon_i(t) \tag{1.8}$$

where the vector of random errors $\boldsymbol{\epsilon}_i = (\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{iJ_i}))$ has a multivariate normal distribution with mean 0 and covariance matrix R_i, and the random errors are assumed to be independent of the random effects. The individual specific mean outcome is $E(Y_i(t)|\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), \boldsymbol{u}_i) = \boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_i$.

To implement a mixed effects model when outcomes are not normally distributed, the outcome distribution needs to be specified. Following this, generalized linear mixed effects models are used to describe the transformed mean of the outcome:

$$g(E(Y_i(t)|\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), \boldsymbol{u}_i)) = \boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_i \tag{1.9}$$

and $E(Y_i(t)|\boldsymbol{X}_i(t), \boldsymbol{Z}_i(t), \boldsymbol{u}_i) = g^{-1}(\boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_i)$. The random effects $\boldsymbol{u}_i$ have some probability distribution with mean 0 and covariance matrix G, and are assumed to be independent of the covariates. Given the random effects, the outcomes are assumed to be independent of each other with $Var(Y_i(t)|\boldsymbol{u}_i) = \phi v(E(Y_i(t)|\boldsymbol{u}_i))$. Maximum likelihood estimation can be used to estimate $\boldsymbol{\beta}$.

The interpretation of regression coefficients in generalized estimating equations can differ when compared to mixed effects models. The coefficients in generalized estimating equation models have a population level interpretation, whereas mixed effects models have an individual level interpretation (however the interpretations are the same for linear and log-linear models). Parameter estimates from the generalized estimating equation approach are consistent even when the within-individual correlation structure is mis-specified given that the mean outcome model has been correctly specified [9]. Mixed effects models require the correct specification of the fixed effects and the correct distributional assumptions for the random effects (and the random errors where appropriate) [10]. Even if the fixed effects are correctly

specified, mis-specifying the random effects distribution can distort estimation of the fixed effects [13]. Furthermore, the sandwich estimator of the covariance structure in generalized estimating equations is robust to mis-specification of the within-individual correlation structure as it provides valid standard errors when the sample size is sufficiently large, while the standard errors in mixed effects models are not robust to mis-specification [14].

1.2.2 Methods for Irregular Longitudinal Data

Generalized estimating equations and mixed effects models can yield consistent regression coefficient estimates when visits are regular (i.e. common set of visiting times across individuals), or completely independent of the outcome process. Generalized estimating equations are also valid when the visit and outcome processes are conditionally independent at time t given whether or not they were censored by time t and fixed covariates. Mixed effects model are valid when the joint likelihood for the outcomes and visits factorizes such that estimation of regression parameters can be based on the outcomes alone. This occurs when the conditional distribution of each waiting time between successive visits given the previous visit times and the outcomes, depends only on previously observed outcomes and on covariates (or a subset of the covariates) [15].

Modelling the Visit Process

When visit times are stochastic, the visit process needs to be explored to ensure valid inference. Recurrent event techniques provide a useful way of characterizing visits. Recurrent event data arise from processes which generate events repeatedly over time [16]. Counting processes provide a useful framework for introducing statistical methods for recurrent event data and describing the visit process. For the i^{th} individual, it is assumed that their corresponding recurrent event process occurs in continuous time starting at time $t = 0$, and that $0 \leq t_{i1} < t_{i2} \ldots$ denote the event times. Let $N_i(t) = \sum_{k=1}^{\infty} I(t_{ik} \leq t)$ represent the number of events occurring by time t (where $I()$ is the indicator function) with $N_i(0) = 0$, and let $N_i(t^+) = \lim_{\Delta t \downarrow 0} N_i(t + \Delta t)$ and $N_i(t^-) = \lim_{\Delta t \downarrow 0} N_i(t - \Delta t)$. The corresponding counting process is $\{N_i(t), 0 \leq t\}$ which is assumed to be right-continuous (i.e. it is assumed that $N_i(t) = N_i(t^+)$). Furthermore, let $H_i(t)$ represent the history of the process at time t. Let $\Delta N_i(t) = N_i(t + \Delta t) - N_i(t)$ and let $dN_i(t) = \lim_{\Delta t \downarrow 0}[N_i(t + \Delta t) - N_i(t)]$.

For the following sections, counting process notation will be used to describe visits. A common feature of many studies is the censoring of individuals, which describes the presence of incomplete information relating to event times. In practice, the censoring of individuals happens for a variety of reasons (e.g. loss to follow-up, administrative censoring etc.). Establishing the reasons leading to censoring can help determine how to handle observed visits and whether it is valid to model visits using a counting process for visit times had there been no censoring. For example, if an individual was lost to follow-up because they moved away, then a counterfactual process would apply under the assumption that they would have continued to visit if they did not move. On the other hand, there are cases where it is debatable whether an individual should be censored. For example, if an individual stopped visiting, then censoring may not necessarily be appropriate because they could have stopped visiting but still remain at-risk for future visits. The decision to use a counterfactual process to model visits depends on the informativeness of censoring. For example, in cases where individuals are at risk of experiencing a terminal event associated with the outcome during the study period (i.e. an event which would halt visits), it can be helpful to

model counterfactual visit times separately from censoring times. To model counterfactual visit times, let $N_i^*(t)$ represent the number of visits for the potentially counterfactual visit times. The modelling of visits is based on the intensity function for recurrent events $\lambda_i(t)$ [17]:

$$\lambda_i(t; H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{Pr(\Delta N_i^*(t) = 1 | H_i(t))}{\Delta t} \tag{1.10}$$

which is analogous to the hazard function in survival analysis. Modelling the visit process in the presence of scheduled visits and unscheduled visits will be discussed in the following section. The main semi-parametric approach is the Andersen-Gill formulation [18] which assumes that events are realizations from a Poisson process with intensity $\lambda_i(t)$ which has a multiplicative structure:

$$\lambda_i(t, \boldsymbol{W}_i(t); \boldsymbol{\alpha}) = \lambda_0(t) \exp(\boldsymbol{W}_i(t)' \boldsymbol{\alpha}) \tag{1.11}$$

where $\boldsymbol{W}_i(t)$ is a r-dimensional vector of covariates, $\boldsymbol{\alpha}$ is a r-dimensional vector of regression coefficients, and $\lambda_0(t)$ is the baseline intensity function at time t. The Andersen-Gill model can be thought of as a Cox proportional hazards model [19] for recurrent events.

Incorporating the Visit Process into the Outcome Model

The following approaches to irregular longitudinal data can accommodate relationships between the visit and outcome process which generalized estimating equations and mixed effects models cannot. Two main semi-parametric approaches to irregular data are inverse-intensity weighted generalized estimating equations and jointly modelling the visit and outcome processes. In both approaches, the visit process is initially modelled using an Andersen-Gill formulation, and then estimation of regression parameters in the outcome model incorporate these results. However, every approach to irregular data relies on a set of assumptions pertaining to the association between the visit and outcome processes which must be evaluated to minimize the potential for biased results.

Choosing the approach to model visits in the presence of scheduled visits and unscheduled visits depends on whether information exists to distinguish scheduled visits from unscheduled visits (e.g. visit tags), and whether scheduled visit times are fixed or stochastic. In cases where scheduled visits cannot be distinguished from unscheduled visits, then all of the visits can be characterized using a single visit process model. In cases where it is known which visits are scheduled or unscheduled, if scheduled visit times are fixed (i.e. no deviations in scheduled visit times), then the unscheduled visit process can be modelled separately. If scheduled visit times are stochastic, then the overall visit process can be modelled as a function of both processes. If visits are irregular (e.g. large deviations in scheduled visit times), then modelling all visits using a single visit process may be appealing. Once it is determined how visits will be modelled, inverse-intensity weighted generalized estimating equations or joint models can be used to model the outcome.

With inverse-intensity weighted generalized estimating equations, visits within individuals are weighted by the inverse of their estimated visit intensity [20]. For a given process $A_i(t)$, let $\bar{A}_i(t) = \{A_i(s), 0 \leq s < t\}$. Applying the notation from Section 1.1 and the counting process notation from above, the observed data by time t are $F^{obs}(t) = \{\bar{N}_i(t), I(C_i \geq t), \bar{\boldsymbol{W}}_i(t), Y_t^{obs}, \bar{\boldsymbol{X}}_i(t), \boldsymbol{W}_i(t)\}$ where $Y_t^{obs} = \{Y_i(s), 0 \leq s < t : \Delta N_i(s) = 1\}$, and $\boldsymbol{W}_i(t)$ represents a vector of auxiliary covariates. Lin *et al* [20] use weighting to

provide consistent estimates of $\boldsymbol{\beta}$ in cases when visits by time t are conditionally independent of the outcome process at time t given the observed data by time t ($F^{obs}(t)$). The weights for subject i are specified as the inverse of their visit intensity which is defined as follows:

$$\lim_{\Delta t \downarrow 0} \frac{Pr(\Delta N_i(t) = 1 | F^{obs}(t), Y_i(t))}{\Delta t} = \lim_{\Delta t \downarrow 0} \frac{Pr(\Delta N_i(t) = 1 | F^{obs}(t))}{\Delta t} = I(C_i \geq t)\lambda_i(t, F^{obs}(t)) \quad (1.12)$$

The marginal outcome model considered is $E(Y_i(t)|\boldsymbol{X}_i(t)) = \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})$, and an Andersen-Gill formulation is assumed for $\lambda_i(t, F^{obs}(t))$. It is assumed that C_i is conditionally independent of the auxiliary covariates and the outcome given $\boldsymbol{X}_i(t)$. The proposed estimating equations from Lin *et al* [20] are:

$$\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{\psi}_i (\boldsymbol{y}_i - \boldsymbol{\mu}_i) = 0 \quad (1.13)$$

where $\boldsymbol{\psi}_i$ is a diagonal weight matrix with the entry of the k^{th} row and k^{th} column being $\frac{s(t_{ik})}{\lambda_i(t_{ik}, F^{obs}(t_{ik}))}$ (for some stabilizing function of time $s(t)$), and $\boldsymbol{V}_i$ is based on a working independence structure. Let $a(t, \boldsymbol{X}(t); \boldsymbol{\beta})$ be a function which is defined as:

$$a(t, \boldsymbol{X}(t); \boldsymbol{\beta}) = \left. \frac{d\mu(t, \boldsymbol{X}(t); \boldsymbol{\beta_0})}{d\boldsymbol{\beta_0}} \right|_{\boldsymbol{\beta_0} = \boldsymbol{\beta}} v(t, \mu(t, \boldsymbol{X}(t); \boldsymbol{\beta_0}))^{-1} s(t) \quad (1.14)$$

where $v()$ denotes the variance function from Section 1.2. Based on this choice of $a(t, \boldsymbol{X}(t); \boldsymbol{\beta})$, the empirical estimating equations can be expressed as:

$$\sum_{i=1}^{n} \int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\hat{\lambda}_i(t, F^{obs}(t))} dN_i(t) = 0 \quad (1.15)$$

The expression $\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\lambda_i(t, F^{obs}(t))} dN_i(t)$ can be shown to have mean 0:

$$\begin{aligned}
&E\left(\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\lambda_i(t, F^{obs}(t))} dN_i(t) \right) \\
&= E\left(E\left[\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\lambda_i(t, F^{obs}(t))} dN_i(t) | F^{obs}(t) \right] \right) \\
&= E\left(E\left[\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\lambda_i(t, F^{obs}(t))} | F^{obs}(t) \right] E\left[dN_i(t) | F^{obs}(t) \right] \right) \\
&\qquad \text{(conditional independence of outcomes and visits given } F^{obs}(t)) \\
&= E\left(E\left[\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) \frac{a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})}{\lambda_i(t, F^{obs}(t))} | F^{obs}(t) \right] I(C_i \geq t) \lambda_i(t, F^{obs}(t)) dt \right) \\
&= E\left(\int_0^{\tau} (Y_i(t) - \mu_i(t, \boldsymbol{X}_i(t); \boldsymbol{\beta})) a(t, \boldsymbol{X}_i(t); \boldsymbol{\beta}) I(C_i \geq t) dt \right) = 0 \\
&\qquad \text{(conditional independence of outcomes and censoring given } \boldsymbol{X}_i(t))
\end{aligned}$$

Buzkova and Lumley [21] recommend using $a(t, \boldsymbol{X}(t); \boldsymbol{\beta}) = \lambda_0(t)$ to ease the computational burden, and thus the weights for each individual visit are $\frac{1}{\exp(\boldsymbol{W}_i(t)' \boldsymbol{\alpha})}$ (where $\boldsymbol{W}_i(t)$ can include functions of $N_i(s), \boldsymbol{X}_i(s)$ and Y_t^{obs} for $s < t$). To implement this procedure in statistical software, $corr(\boldsymbol{y}_i)$ must

have an independence structure to ensure that weights are correctly incorporated into analyses on the outcome. Most statistical software divides the scale parameter ϕ by the inputted weights, and thus the correct estimating equations will be used only if $corr(\boldsymbol{y}_i)$ has an independence structure.

With joint models, the dependence between the visit and outcome processes is captured using shared random effects. Joint models can accommodate certain relationships between the visit process and outcome process which inverse-intensity weighted generalized estimating equations cannot. Let u_{i1} denote the random effect for the outcome model (and let $\boldsymbol{u}_{i1}$ be a q-dimensional vector), let u_{i2} denote the random effect for the visit process model (if necessary, let u_{is} denote a random effect that is the same in the outcome model and visit process model), and let u_{i3} denote the random effect for the censoring model. There are many formulations of semi-parametric joint models which differ in terms of the structure of the outcome and visit process models (e.g. additive or multiplicative), and restrictions on covariates in both models (e.g. time-varying or fixed) (see Table 1.1).

Proposed Formulation	**Model for Mean Outcome**	**Model for Visit Process**	**Model for Censoring**
Liang *et al* [22]	$\gamma_0(t) + \boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_{i1}$	$u_{i2}\lambda_0(t)\exp(\boldsymbol{W}_i'\boldsymbol{\alpha})$	None
Sun *et al* [23]	$u_{is}\gamma_0(t)\exp(\boldsymbol{X}_i'\boldsymbol{\beta})$	$u_{is}\lambda_0(t)\exp(\boldsymbol{X}_i'\boldsymbol{\alpha})$	None
Song *et al* [24]	$\gamma_0(t) + \boldsymbol{X}_i(t)'\boldsymbol{\beta} + u_{i1}$	$u_{i2}\lambda_0(t)\exp(\boldsymbol{X}_i(t)'\boldsymbol{\alpha})$	None
Sun *et al* [25]	$\gamma_0(t; u_{i1}) + \boldsymbol{X}_i'\boldsymbol{\beta}$	$\lambda_0(t; u_{i2})\exp(\boldsymbol{X}_i'\boldsymbol{\alpha})$	None
Su *et al* [26]	$\gamma_0(t; u_{i1}) + \boldsymbol{X}_i'\boldsymbol{\beta}$	$\lambda_0(t; u_{i2})\exp(\boldsymbol{X}_i'\boldsymbol{\alpha})$	$h_0(t, u_{i3}) + \boldsymbol{X}_i'\boldsymbol{\zeta}$

Table 1.1: Semi-parametric joint models for the visit and outcome processes where $\gamma_0(t)$ is an unspecified smooth function of time in the outcome model, $\boldsymbol{X}_i(t)$ is a vector of covariates by time t, $\boldsymbol{\beta}$ is a vector of regression coefficients for the outcome model, $\boldsymbol{Z}_i(t)$ is a vector of covariates by time t for the random effects in the outcome model, u_{is}, u_{i1}, $\boldsymbol{u}_{i1}$, u_{i2} and u_{i3} are random effects, $\boldsymbol{W}_i$ and $\boldsymbol{X}_i$ are vectors of fixed covariates, $\boldsymbol{\alpha}$ is a vector of regression coefficients for the visit process model, $\lambda_0(t)$ is the baseline intensity function by time t, $h_0(t)$ is the baseline hazard function by time t for the censoring model, and $\boldsymbol{\zeta}$ is a vector of regression coefficients for the censoring model.

These methods assume that the outcome and visit process are conditionally independent given the random effects and covariates.

Similar to inverse-intensity weighted generalized estimating equations, joint models are also based on estimating equations with mean 0. The formulation of Liang *et al* [22] will be used to demonstrate the estimation procedure behind joint models. Liang *et al* [22] assume that u_{i2} is non-negative with mean 1 and variance σ^2, that $E(\boldsymbol{u}_{i1}|u_{i2}) = \boldsymbol{\theta}(u_{i2} - 1)$ where $\boldsymbol{\theta}$ is a vector of scale parameters, and that censoring is independent of visit times and the outcomes given the covariates. To illustrate why $\boldsymbol{\theta}$ is identifiable, consider the case where there is only 1 scale parameter (i.e. θ). If individuals with higher rates of visits were also observed to have higher outcomes, then the estimate of θ would be positive (and if they were observed to have lower outcomes, then the estimate of θ would be negative). However, the vector of scale parameters $\boldsymbol{\theta}$ is not identifiable if $\sigma^2 = 0$.

The conditional distribution of the total number of visits J_i given u_{i2} is Poisson with mean $u_{i2}\Lambda_0(C_i)\exp(\boldsymbol{W}_i'\boldsymbol{\alpha})$ where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the baseline cumulative intensity function by time t. Given J_i, C_i, and u_{i2}, the visit times $t_{i1}, \ldots, t_{iJ_i}$ are order statistics with the following density

function:

$$P(t_{i1},\ldots,t_{iJ_i}|J_i,u_{i2},C_i) = J_i!\prod_{j=1}^{J_i}\frac{d\Lambda_0(t_{ij})}{\Lambda_0(C_i)} \tag{1.16}$$

Liang *et al* [22] specify the following linear outcome model:

$$Y_i(t) = \gamma_0(t) + \boldsymbol{X}_i(t)'\boldsymbol{\beta} + \boldsymbol{Z}_i(t)'\boldsymbol{u}_{i1} + \epsilon_i(t) \tag{1.17}$$

and using the following result:

$$\begin{aligned} E(dN_i(t)|J_i,C_i) &= E(dN_i(t)|J_i,C_i,u_{i2}) \\ &= I(C_i \geq t)J_i\frac{d\Lambda_0(t)}{\Lambda_0(C_i)} \end{aligned}$$

it follows that:

$$\begin{aligned} &E[(Y_i(t) - \boldsymbol{X}_i(t)'\boldsymbol{\beta})dN_i(t)|J_i,C_i] \\ =&E[E(\gamma_0(t)dN_i(t) + \boldsymbol{Z}_i(t)'\boldsymbol{u}_{i1}dN_i(t)|J_i,C_i,u_{i2})|J_i,C_i] \\ =&E[\gamma_0(t)E(dN_i(t)|J_i,C_i,u_{i2}) + \boldsymbol{Z}_i(t)'E(\boldsymbol{u}_{i1}dN_i(t)|J_i,C_i,u_{i2})|J_i,C_i] \\ =&I(C_i \geq t)J_i\frac{d\Lambda_0(t)}{\Lambda_0(C_i)}\Big(\gamma_0(t) + \boldsymbol{Z}_i(t)'E[E(\boldsymbol{u}_{i1}|J_i,C_i,u_{i2})|J_i,C_i]\Big) \\ =&I(C_i \geq t)J_i\frac{d\Lambda_0(t)}{\Lambda_0(C_i)}\Big(\gamma_0(t) + \boldsymbol{Z}_i(t)'\boldsymbol{\theta}E[u_{i2}-1|J_i,C_i]\Big) \end{aligned}$$

and using this result, it becomes apparent that:

$$\begin{aligned} &E[(Y_i(t) - \boldsymbol{X}_i(t)'\boldsymbol{\beta} - \boldsymbol{Z}_i(t)'\boldsymbol{\theta}E[u_{i2}-1|J_i,C_i])dN_i(t)|J_i,C_i] \\ =&E[(Y_i(t) - \boldsymbol{X}_i(t)'\boldsymbol{\beta} - \boldsymbol{Z}_i(t)'\boldsymbol{\theta}E[u_{i2}-1|J_i,C_i])|J_i,C_i]E[dN_i(t)|J_i,C_i] \\ =&\gamma_0(t)E[dN_i(t)|J_i,C_i] \\ =&\gamma_0(t)I(C_i \geq t)J_i\frac{d\Lambda_0(t)}{\Lambda_0(C_i)} \end{aligned}$$

Letting $A(t) = \int_0^t \gamma_0(s)d\Lambda(s)$, the expression:

$$M_i(t) = \int_0^t (Y_i(s) - \boldsymbol{X}_i(s)'\boldsymbol{\beta} - \boldsymbol{Z}_i(s)'\boldsymbol{\theta}E[u_{i2}-1|J_i,C_i])dN_i(s) - I(C_i \geq s)J_i\frac{dA(s)}{\Lambda(C_i)} \tag{1.18}$$

has mean 0 when $\Lambda = \Lambda_0$, and therefore $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ can be estimated by solving the following two estimating equations:

$$\begin{aligned} &\sum_{i=1}^n dM_i(t) = 0 \\ &\sum_{i=1}^n \int_0^\tau \begin{pmatrix}\boldsymbol{X}_i(t) \\ \hat{\boldsymbol{B}}_i(t)\end{pmatrix} dM_i(t) = 0 \end{aligned} \tag{1.19}$$

Methods for irregular longitudinal data can accommodate a broader range of visiting scenarios when compared to standard mixed effects models and generalized estimating equations. However, choosing an appropriate method requires careful consideration of the visit process.

1.3 Exploring Irregularity

Modelling the visit process can help evaluate the appropriateness of a method for irregular longitudinal data. In missing data, predictors of missingness can be assessed using a logistic regression model to help distinguish between data missing at random and missing completely at random. In a longitudinal clinical trial comparing rates of amenorrhea between contracepting women who were assigned 100mg or 150mg of depot-medroxyprogesterone acetate (DMPA) [27], a logistic regression model indicated that missingness was associated with whether or not amenorrhea occurred at the previous visit (this effect was not significantly different across the two dose groups). The data were deemed to be missing at random and thus inverse-probability weighted generalized estimating equations were considered. However, it is not possible to empirically determine whether data are missing at random or missing not at random. In cases when assumptions about the missing data mechanism are not testable, Scharfstein *et al* [28] recommend using sensitivity analyses to assess changes in inference.

With irregular data, a visit process model can be used to identify predictors of visit intensity and help determine whether the visit process is conditionally independent of the outcome process given the observed data ($F^{obs}(t)$) (assuming there are no unmeasured variables). For example, Lin *et al* [20] apply inverse-intensity weighted generalized estimating equations to a study which modelled the levels of homelessness in the past 3 months amongst individuals with mental illness. The visit process model indicated that the effects of recorded income from the past 3 months, quality of living, and whether participants received social benefits on visit intensity significantly differed across the three treatment groups (standard care, case-manager only, voucher and case-manager). This analysis provided a justification for weighting generalized estimating equations by the inverse of the estimated visit intensity. Although regression models can help evaluate certain visit process assumptions, it can be difficult to determine at what point the extent of irregularity merits specialized methods for irregular longitudinal data.

It would be ideal to have a measure which can be used to quantify the extent of irregularity and to help decide whether methods for repeated measures are appropriate. There are several examples of measures that are used to inform analysis or interpretations of results in other areas of statistics. For example, meta-analysis has the tau-squared estimator [29] and the I-squared statistic [30]. The tau-squared estimator in a random effects meta-analysis describes the extent of variation of the effect size parameters across different studies with larger values suggesting that a fixed effects model may not be appropriate. However, the tau-squared can only be compared across meta-analyses when the outcomes have the same definition, and interpreting an individual tau-squared estimate can be difficult depending on how the effect is measured (e.g. on the log-odds scale). The I-squared statistic was developed to overcome the limitations of the tau-squared estimator and other measures. The I-squared statistic describes the percentage of variability across studies which can be attributed to heterogeneity with larger values of the I-squared statistic suggesting increasing heterogeneity. When the extent of heterogeneity is large, researchers would explore reasons for the heterogeneity and consider a random effects approach [29]. In addition to summarizing the extent of heterogeneity, the I-squared statistic was intended to be straightforward to interpret, and not depend on the type of outcome (e.g. in meta-

analyses where some studies have a dichotomous outcome and some studies have a continuous outcome), or the number of studies.

Another example of a measure is the variance inflation factor (VIF) [31]. The VIF for the k^{th} predictor in a multiple linear regression model is calculated as $\frac{1}{1-R_k^2}$ where R_k^2 is obtained from regressing the k^{th} predictor on the other predictors in the multiple linear regression model. The VIF reports the amount of additional variation in a model term which is due to multi-collinearity. The smallest value the VIF can be is 1 which implies no multi-collinearity. In practice, a VIF between 1 and 4 is taken to indicate a moderate level of multi-collinearity but is deemed acceptable as some level of multi-collinearity may be unavoidable [32]. On the other hand, a VIF above 4 is taken to indicate a problematic level of multi-collinearity requiring some predictors to be discarded from the model [32]. The VIF was meant to be an intuitive index of the effects of multi-collinearity on the variance of a model term.

Drawing on the fact that the concept of irregular data is analogous to missing data, the proportion of individuals with missing data is another example of an intuitive measure which can guide researchers in selecting a modelling approach. Depending on the suspected missing data mechanism, different values of missingness can present a cause for concern [33], and even deter further analyses. This is also the case with irregular data as depending on the informativeness of the visit process, the extent of irregularity can also be problematic [15]. However, measures for quantifying the extent of irregularity are lacking [34]. The purpose of this thesis is to propose measures for quantifying the extent of irregularity.

1.4 Thesis Objectives and Structure

The main objective of this thesis is to propose measures for quantifying the extent of visit irregularity in longitudinal data which are intuitive, and invariant to sample size. In addition, these measures are intended to help make the decision of whether data should be treated as repeated measures, or whether methods for irregular longitudinal data should be utilized. Due to the novelty of these measures, this thesis aims to establish properties of these measures under a range of visit scenarios.

Objective 1: The first objective is to formulate visual measures for summarizing the extent of irregularity.

Approach: The measures of irregularity will be based on bins defined across the study period. We will estimate the proportions of individuals with 0, 1, and >1 visits per bin. When the choice of bin widths is not obvious, we vary bin widths and plot the mean proportions of individuals with 0, 1, and >1 visits per bin. With perfect repeated measures, the mean proportions of individuals with 1 visit per bin are 1, and the mean proportions of individuals with 0 and >1 visits per bin are 0. Therefore, the estimated mean proportions will be compared to values that are consistent with perfect repeated measures.

Objective 2: The second objective is to develop a single numerical measure for quantifying the extent of irregularity that increases with increasing irregularity.

Approach: We propose plotting the mean proportions of individuals with 0 visits per bin against the mean proportions of individuals with >1 visit per bin (as bin width is varied), and using the area under the curve (AUC) as a single measure for quantifying the extent of irregularity. When there are

pre-specified visit times, the AUC will be compared to 0 which corresponds to an AUC from perfect repeated measures. To overcome the influence of missing data on the AUC, a likelihood-based AUC assuming no missingness will be derived. To form a comprehensive assessment of irregularity, the AUC will be interpreted in conjunction with the mean proportions of individuals with 0, 1, and >1 visits per bin.

This is a sandwich thesis. In Chapters 2 and 3 respectively, I address objectives 1 and 2 above. In Chapter 4, I provide a comprehensive guide for summarizing the extent of irregularity based on the proposed measures from Chapters 2 and 3, and discuss how to use the information obtained from these measures to inform the modelling approach for the outcome. In Chapter 5, I perform a complete demonstration of how to handle irregular longitudinal data by applying the measures of this thesis to a longitudinal study of depression, and then analyze the outcome using the appropriate modelling approach. Chapter 6 contains a discussion of the main findings including possible avenues for future research. A section with R code along with a worked example appears in Appendix C.

1.5 Author's Contributions

For Chapter 2, these are the author contribution statements:

- **Armend Lokku:** Conceptualized and operationalized the measures, performed all statistical analyses, interpreted results, and drafted, critically reviewed, and approved the final manuscript.
- **Eleanor Pullenayegum:** Helped conceptualize the measures and plan statistical analyses, critically reviewed and approved the final manuscript.
- **Lily Lim:** Coordinated and supervised data collection for the cSLE study, interpreted results, and critically reviewed and approved the final manuscript.
- **Catherine Birken:** Coordinated and supervised data collection for the TARGet Kids! data, interpreted results, and critically reviewed and approved the manuscript.

For Chapter 3, these are the author contribution statements:

- **Armend Lokku:** Formulated the AUC, developed likelihood-based estimation of the AUC, designed and implemented all statistical analyses, interpreted results, and drafted, critically reviewed, and approved the final manuscript.
- **Eleanor Pullenayegum:** Helped conceptualize the measures and plan statistical analyses, critically reviewed and approved the final manuscript.
- **Catherine Birken:** Coordinated and supervised data collection for the TARGet Kids! data, interpreted results, and critically reviewed and approved the manuscript.
- **Jonathon Maguire:** Coordinated and supervised data collection for the TARGet Kids! data, interpreted results, and critically reviewed and approved the manuscript.

The manuscripts from Chapter 2 and Chapter 3 are:

- A. Lokku, L. SH. Lim, C. S. Birken, and E. M. Pullenayegum. Summarizing the extent of visit irregularity in longitudinal data. *BMC Medical Research Methodology*, 20:135, 2020.
- A. Lokku, C. S. Birken, J. L. Maguire, and E. M. Pullenayegum. Quantifying the extent of visit irregularity in longitudinal data. *In submission.*

Chapter 2

Summarizing the Extent of Visit Irregularity in Longitudinal Data

Abstract

Background: Observational longitudinal data often feature irregular, informative visit times. We propose descriptive measures to quantify the extent of irregularity to select an appropriate analytic outcome approach.

Methods: We divided the study period into bins and calculated the mean proportions of individuals with 0, 1, and >1 visits per bin. Perfect repeated measures features everyone with 1 visit per bin. Missingness leads to individuals with 0 visits per bin while irregularity leads to individuals with >1 visit per bin. We applied these methods to: 1) the TARGet Kids! study, which invites participation at ages 2, 4, 6, 9, 12, 15, 18, 24 months, and 2) the childhood-onset Systemic Lupus Erythematosus (cSLE) study which recommended at least 1 visit every 6 months.

Results: The mean proportions of 0 and >1 visits per bin were above 0.67 and below 0.03 respectively in the TARGet Kids! study, suggesting repeated measures with missingness. For the cSLE study, bin widths of 6 months yielded mean proportions of 1 and >1 visits per bin of 0.39, suggesting irregular visits.

Conclusions: Our methods describe the extent of irregularity and help distinguish between protocol-driven visits and irregular visits. This is an important step in choosing an analytic strategy for the outcome.

2.1 Background

Observational longitudinal data often feature visit times that vary across individuals with the potential for the timings and frequency of visits to be associated with the study outcome. Visit irregularity can lead to misleading conclusions [20] and should therefore be accounted for in analyses of the outcome trajectory [21]. For example, in a randomized trial of the interventions to reduce homelessness, individuals with greater levels of homelessness were likely to visit more frequently [20]. When the visit process was ignored the group receiving a case manager only had 0.71% more days homeless than the standard care; when the visit process was accounted for the effect estimate reversed direction with the case manager group having 1.64% fewer days homeless. In another example, Buzkova *et al* [35] estimated the prevalence of pneumonia amongst Kenyan mothers with HIV-1 to be 2.89% when the visit process was ignored; the estimate almost halved to 1.48% after accounting for visits. Observational data are readily available (e.g. in administrative databases, electronic medical records); however, data collected over the course of usual care are particularly liable to irregular visit patterns.

The problem of visit irregularity is analogous to missing data. The key difference between irregular data and missing data is that the latter occurs when a scheduled measurement is not recorded, whereas irregular data describes the presence of imbalanced visit patterns across individuals, often in the absence of a study wide follow-up schedule. In statistical terms, data is missing when visit times are fixed by design and whether the visit occurs is a random variable. With irregular visits, the timings of visits are the random variables.

The possibility for biased results in the presence of missing data is generally recognized in applied settings [36], and this consensus has led to the exploration of missing data patterns being recommended (e.g. STROBE, CONSORT 2010) [37] [38]. Summarizing missing data typically begins by recording the frequency (or percentage) of individuals with missing values for each variable (STROBE [37]), upon which the severity of the problem can be judged. For example, if the data is judged to be missing at random (or completely at random), one might proceed with techniques that deal with missing longitudinal data such as multiple imputation [39] or inverse-probability weighting [40]. On the other hand, unless missingness is known to be completely at random, missing values may render further analysis futile as informative missingness can lead to bias as missingness increases.

Given that irregularity can lead to bias, irregular data should be explored with the same rigor as is done with missing data. Irregularity exists on a continuum where on one extreme the extent of irregularity can vary to the point where no two individuals share the same visit times. At the other extreme, visit times can resemble perfect repeated measures where every individual has 1 visit at each pre-specified visit time in the protocol. In practice, there are scenarios between both extremes where visits are intended to be repeated measures but the timings of scheduled visits vary across individuals, scheduled visits are missing, or there are unscheduled visits. There are different techniques for analyzing irregular data versus repeated measures, but it can be difficult to decide at what point the data should no longer be treated as repeated measures, but as irregular data. Farzanfar *et al* [41] performed a systematic review of longitudinal studies to explore how irregularity is reported and handled in practice. They observed that of the 44 eligible studies: 86% of the studies did not report enough information to assess if it was necessary to account for informative visit timings, 3 studies reported on the gaps between visits, 2 studies assessed predictors of visit times, and only 1 study used a specialized method for irregular longitudinal data. One of the reasons why visit irregularity is ignored in practice is that

most of the literature on this topic is highly technical.

There are currently no proposed measures for quantifying the extent of irregularity in longitudinal data. This paper provides intuitive visual measures that can be used by researchers who are not experts in statistics along with the respective R code to implement these measures in practice. This paper demonstrates how these descriptive measures can help distinguish between individually-driven irregular visits and protocol-driven regular visits, and illustrates how to examine the underlying visit process to select an appropriate statistical approach for the outcome.

2.2 Methods

2.2.1 Datasets

We will illustrate our proposed measures of irregularity with the following two datasets.

Pre-Specified Visit Times: TARGet Kids!

The TARGet Kids! study enrolls healthy children aged 0-5 years and follows them until age 18, with the aim of investigating the relationship between early life exposures and later health problems including obesity, micronutrient deficiencies, and developmental problems [8]. Well-child visits are recommended at ages 2, 4, 6, 9, 12, 15, 18, 24 months, and then every 12 months afterwards, with vaccinations occurring at ages 2, 4, 6, 12, 15, 18 months. Parents also bring their children for "sick" visits as needed. Individuals are recruited and enrolled by research assistants who approach them at well-child visits. In general, most well-child visits did not occur prior to the expected visit schedule because the physician could not bill for an early visit, and vaccinations could only occur once a child reaches a specific age. For example, the Measles, Mumps and Rubella (MMR) vaccine cannot be administered until a child is 12 months old.

No Pre-Specified Visit Times: Child Systemic Lupus Erythematosus Study

The child lupus study was a retrospective inception cohort study of patients who were diagnosed with childhood-onset Systemic Lupus Erythematosus (cSLE) and followed at a single center with a dedicated cSLE clinic. This cohort was followed from childhood into adulthood. Visit dates ranged from January 1st, 1985 to September 30th, 2011. Individuals were followed at least once every 6 months; however, visit frequency depended on the severity of their disease. The primary objective of this study was to assess differences in disease activity trajectories among all cSLE patients.

2.2.2 Measures for Quantifying the Extent of Visit Irregularity

The following measures can be used to assess the extent of visit irregularity and help inform the modelling approach for the outcome. They can also help determine whether observed visits can be viewed as repeated measures subject to missingness. The proposed measures are based on techniques used to explore missing data. In a repeated measures design, summarizing missing values begins by recording the percentage of missing values at each pre-specified visit time. In addition, predictors of being observed at a pre-specified visit time can be evaluated using a regression model (e.g. logistic regression). We adapt

these concepts to the context of irregular data. We consider studies with pre-specified visit times in the protocol, and studies which do not pre-specify visit times in the protocol.

Pre-Specified Visit Times

We propose constructing bins around pre-specified visit times. Let the time frame of interest be $(0, \tau)$, and let T_j denote the j^{th} pre-specified visit time ($j = 1, 2, \ldots k$). The j^{th} bin is given by the interval (L_j, R_j), where L_j and R_j are chosen to specify the left and right cut-points of the j^{th} bin respectively (Figure 2.1).

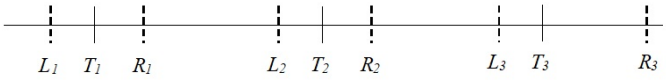

Figure 2.1: Specifying bin widths for pre-specified visit times.

We require that $R_j < L_{j+1}$ (for all values of j) so that bins do not overlap, and that $L_j < T_j < R_j$. These bins can be used to calculate summary statistics such as the proportions of individuals with 0, 1, and >1 visits per bin.

Bin widths should be specified according to clinical context as appropriate. For example, the HbA1C blood test measures blood glucose levels from the previous 3 months (levels are known to be stable during this time period [42]), and hence bin widths should not be less than this. Bins can have different widths across the study period to account for known patterns in visit intensity (e.g. more frequent visits in the winter). Another approach to specifying bin widths is to use the percentage of the time gap between the pre-specified visit times (T_j). For example, 10% of the gap implies that $L_j = T_j - 0.1(T_j - T_{j-1})$ and $R_j = T_j + 0.1(T_{j+1} - T_j)$. When there is no obvious choice of bin widths, reporting on varying bin widths can be helpful.

In perfect repeated measures, all individuals have 1 visit in a bin (regardless of bin width) and no individuals have 0 or >1 visits per bin. Thus the proportion of individuals with 0 or >1 visits per bin are 0 and the proportion of individuals with 1 visit per bin is 1. Figure 2.2 illustrates the visit timings for a random subset of 20 individuals from a perfect repeated measures simulated dataset with 100 observations and five pre-specified visit times (2, 4, 6, 8, 10 months).

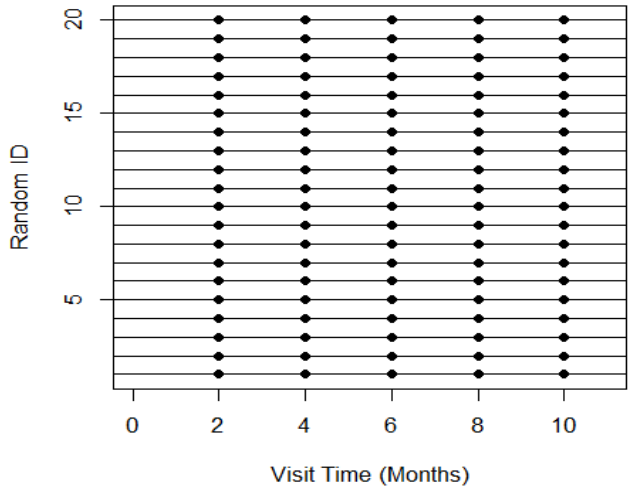

Figure 2.2: The visit timings for a random subset of 20 individuals from a perfect repeated measures simulated dataset with 100 observations.

As the levels of missingness increase, the proportion of individuals with 0 visits per bin increases. As irregularity increases, the proportion of individuals with >1 visit per bin increases.

The R code for plotting visit patterns for a random subset of individuals and the mean proportions of individuals with 0, 1, and >1 visits per bin uses the "IrregLong" package in CRAN [43] and is presented in Appendix A.

No Pre-Specified Visit Times

We construct adjacent bins across the entire study period. Bin widths can be determined by clinical context or known visit patterns (e.g. fewer visits later on in follow-up could be accommodated by wider bins). The j^{th} bin is given by the interval (L_j, R_j), where L_j and R_j are chosen to specify the left boundary and right boundary of the j^{th} bin respectively (Figure 2.3).

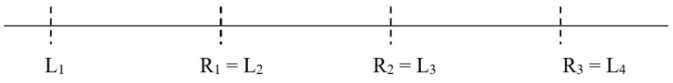

Figure 2.3: Specifying bin widths for no pre-specified visit times.

The mean proportions of individuals with 0, 1, and >1 visits per bin can be obtained by varying the number of bins (as the number of bins increases, bin widths decrease). These values can be used to judge the extent of irregularity by assessing whether or not they are consistent with values that would result from repeated measures. The larger the disparity of these values from repeated measures values suggests the greater the extent of irregularity. To evaluate this, the first step is to plot the mean proportions of individuals with 0, 1, and >1 visits per bin as a function of bin width. The next step is to identify the bin width that yields the largest mean proportion of individuals with 1 visit per bin (i.e. in perfect

repeated measures, all individuals have 1 visit per bin). At this bin width, determine if either the mean proportions of individuals with 0 or >1 visits per bin are 0. If the mean proportion of individuals with >1 visit per bin is not 0, this indicates a degree of irregularity. If the mean proportion of individuals with >1 visit per bin is 0 and the mean proportion of individuals with 0 visits per bin is not 0, this suggests the data can be viewed as repeated measures with missingness. This comparison can be supplemented by identifying the largest bin width such that the mean proportion of individuals with >1 visit per bin is 0, and evaluating whether the mean proportions of individuals with 0 and 1 visits per bin are 0. If at the largest such bin width, the mean proportion of individuals with 0 visits per bin is 0 and the mean proportion of individuals with 1 visit per bin is not 0, this suggests the data can be treated as repeated measures.

Censoring

Both left and right censoring should be considered when using bins to explore visit irregularity. Individuals may enter the study after the first pre-specified visit time, and the dataset may be closed before they have the opportunity to attend all the follow-up visits. In cases where censoring is administrative and unlikely to lead to bias, we may wish to measure irregularity separately from censoring. This can be done by specifying an "at-risk" set of individuals for each bin (i.e. individuals who are under follow-up for all times in the bin) then using just these individuals to estimate the proportions of 0, 1, and >1 visits per bin. Individuals who are lost to follow-up (rather than administratively censored) can still be at-risk beyond their last visit. However, individuals should not be considered in calculations for bins representing times before they entered the study or after the dataset was closed.

2.3 Results

2.3.1 Pre-Specified Visit Times: TARGet Kids!

The study comprised of 6,470 individuals with a median follow-up of 5.32 years. The years of recruitment ranged from 2008 to 2015. Data from well-child visits and sick visits were used to assess whether the data resembled repeated measures. Visits from all 6,470 individuals were included in bin calculations, and Figure 2.4 displays the age at each visit for a random subset of 20 individuals.

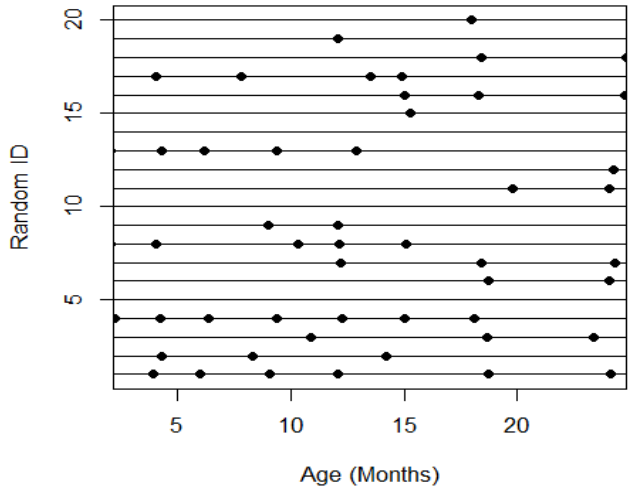

Figure 2.4: The age at visit (months) for a random subset of 20 individuals from the TARGet Kids! cohort.

All bins were anchored on the ages of well-child visits and the left side of each bin was fixed at 5% of the gap between successive well-child visit ages (since visits could not occur too early) and the right side of each bin was varied from 1% to 95% of the gap. Figure 2.5 illustrates the mean proportions of individuals with 0, 1, and >1 visits per bin across varying bin widths.

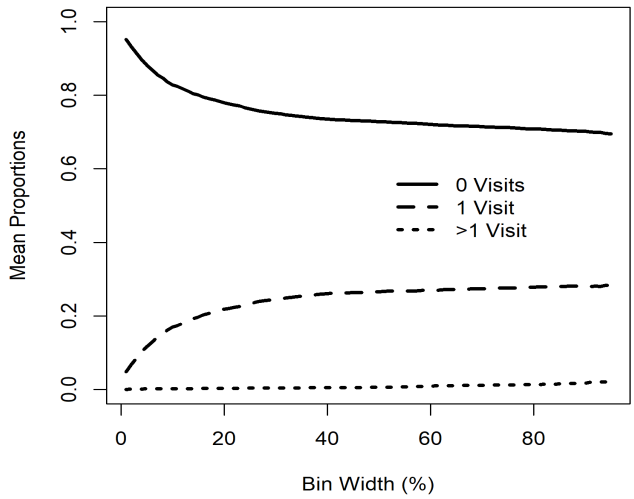

Figure 2.5: The mean proportions of individuals with 0, 1, and >1 visits per bin as bin width varies from 1% to 95% of the gap between well-child visits for the TARGet Kids! cohort.

The mean proportions of individuals with 0 visits per bin were above 0.67 for all bin widths while

the mean proportions of individuals with >1 visit per bin were below 0.03. These values suggest that individuals mostly visit according to suggested visit times. The pattern is similar to repeated measures subject to missingness.

2.3.2 No Pre-Specified Visit Times: cSLE Study

The study size was 473 individuals with a median duration of follow-up of 5.44 years (total duration of follow-up was 2,666 patient-years). Figure 2.6 illustrates visit timings for a random subset of 20 individuals.

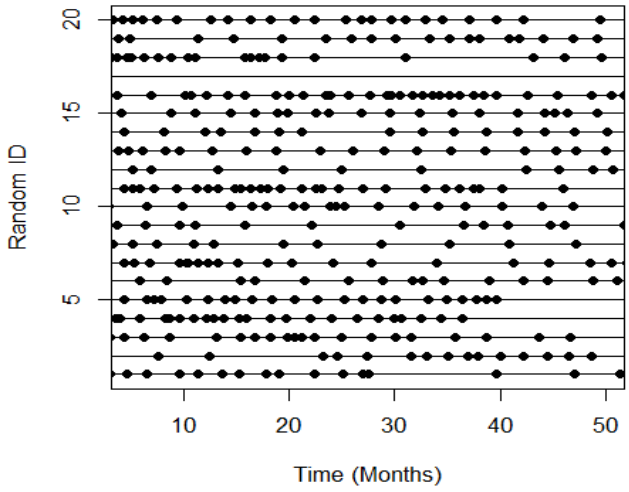

Figure 2.6: The visit timings for a random subset of 20 individuals from the cSLE study.

Visit schedules highly varied and were personalized with few individuals having similar visit patterns.

To determine the extent of visit irregularity, the entire study period was split into adjacent and equally-sized bins and the number of bins was varied. Figure 2.7 shows the mean proportions of individuals with 0, 1, and >1 visits per bin across bin widths.

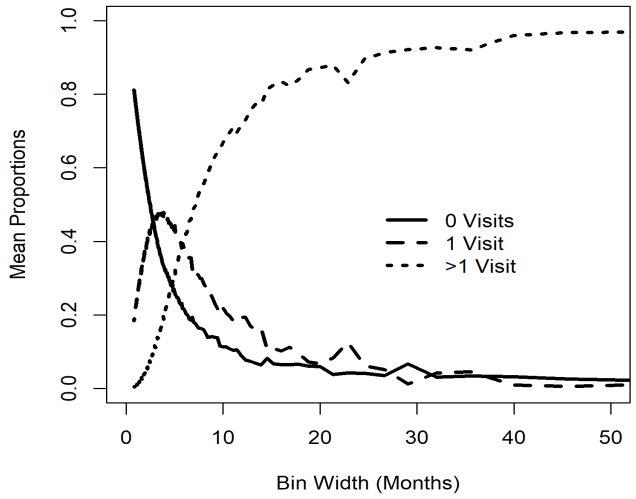

Figure 2.7: The mean proportions of individuals with 0, 1, and >1 visits per bin for the cSLE study.

When the disease is controlled, individuals are recommended to visit every 6 months, and if their disease status worsens they visit more frequently. For bin widths of 6 months, the mean proportion of individuals with >1 visit per bin was 0.39, the mean proportion of individuals with 1 visit per bin was 0.39, and the mean proportion of individuals with 0 visits per bin was 0.22. Although individuals were expected to visit at least once every 6 months, 22% of individuals on average had 0 visits when using this interval. The mean proportions of individuals with 1 visit per bin had a maximum value of 0.48 corresponding to bin widths of 3.52 months (the mean proportion of individuals with >1 visit per bin was 0.15, and the mean proportion of individuals with 0 visits per bin was 0.37). For smaller bin widths of 0.82 months, the mean proportion of individuals with >1 visit per bin was 0.004, the mean proportion of individuals with 0 visits per bin was 0.81, and the mean proportion of individuals with 1 visit per bin was 0.19. There were no bin widths that were consistent with repeated measures because even when the mean proportions of individuals with 1 visit per bin was maximized, 52% of individuals on average had >1 or 0 visits per bin, and when bin widths were small enough such that the mean proportion of individuals with >1 visit per bin was almost 0, 82% of individuals on average did not contribute data because they had 0 visits per bin. These results suggest individually-driven irregular visits, and therefore the extent of irregularity needs to be considered in analyses of the disease trajectory.

2.4 Analyzing Irregular Visit Processes

There are several methods which can accommodate irregular visit processes, but they make assumptions concerning the relationship between the outcome and irregularity [15]. It is important to consider the irregularity mechanism to ensure the validity of any chosen statistical approach for the outcome.

With missing data, judging whether data are missing completely at random (MCAR) or missing at random (MAR) is done by evaluating predictors of being observed at pre-specified occasions. This is typically done by comparing demographic and other available characteristics across the observed and

unobserved groups using tables or logistic regression models [44]. With irregular data, the relationship between the outcome and visit process can be judged by identifying predictors of visit intensity.

2.4.1 Visit Processes

Determining the visit process is important because all methods make assumptions concerning the relationship between the visit and the outcome processes. Visit processes can be regular or irregular, and among irregular processes, the taxonomy for classifying missing data mechanisms has been extended to irregular visit processes [15]:

a) **Visiting completely at random (VCAR):** Visit times are completely independent of the outcome process.

b) **Visiting at random (VAR):** Given the observed history (outcomes, visits, covariates) up to time t, the visit process at time t, is independent of the outcome process.

c) **Visiting not at random (VNAR):** Given the observed history (outcomes, visits, covariates) up to time t, the visit process is **not** independent of the outcome process at time t.

This classification scheme highlights the potential relationships between the outcome and visit processes over time and can be used to determine the appropriate analytic method for the outcome [15].

2.4.2 Determining the Visit Process

To determine the visit process, it can be helpful to consider the study protocol. Some protocols prespecify a common set of visit times (fixed visits), while others allow current patient status to determine future visit times such as: 1) a patient's previously observed history (history-dependent), 2) physician-driven visits, or 3) self-determined or patient-driven visits.

If the protocol is adhered to perfectly, then history-dependent visits correspond to VAR. Physician-driven visits can also result in VAR provided that all the information that the physician uses to decide the time of the next visit is recorded in the patient's chart. Patient-driven visits may be VNAR because the underlying factors which influence future visits are usually not reported in advance. It is important to consider the extent of deviations from pre-specified visit times for fixed, history-dependent protocols and physician-driven visits because the visit process may be non-ignorable, especially if deviations are due to unobserved or unrecorded factors.

Although it is possible to distinguish between VAR and VCAR visits using recurrent event regression models, there is no way of distinguishing between VAR and VNAR visits. Any modelling assumptions should be judged carefully to avoid biased results on the outcome.

Distinguishing between VCAR and VAR: Modelling the Visit Process

Identifying predictors of visit intensity can be performed using recurrent event regression models. Techniques for analyzing recurrent event data are well established [16] and are applicable to irregular visits. Regression models for recurrent events characterize event rates over time by modelling the intensity function [17]. The intensity function is analogous to the hazard function in survival analysis in the sense that it can be thought of as the instantaneous probability of observing an event by time t, conditional on a subject's observed history.

One of the more commonly used intensity regression models is the Andersen-Gill model [18], which is an extension of the Cox proportional-hazards regression model [19]. The Andersen-Gill model is quite flexible as it can include time-dependent factors and past observed outcomes as predictors of future event intensity. The Andersen-Gill model can be implemented in standard survival analysis software such as R 3.1.0 [45].

Application to the cSLE Study

Exploration of visits using bins indicated irregularity, therefore the visit process must be addressed. The following analyses aimed to identify predictors to help distinguish between VCAR and VAR. This was done by fitting a Cox proportional hazards regression model using the Andersen-Gill formulation with age at visit as the time variable. Baseline characteristics included: age at diagnosis (years), sex, race (Caucasian, Black, Asian, and Other), number of American College of Rheumatology (ACR) criteria for SLE at diagnosis, the presence of lupus nephritis at baseline, and mortality. Time-varying predictors included: disease activity measured by the SLE disease activity index [46] [47], prednisone dose, anti-malarial medication, total organ damage as measured by the SLE damage index [46], bone damage, cardiovascular damage (acute myocardial infarction, cerebrovascular accidents, and myocardial failure), a composite score for use of significant immunosuppression (any use of azathioprine for major organ disease, cyclophosphamide, cyclosporine, tacrolimus), and major organ involvement (including cerebrovascular accidents, psychosis, lupus nephritis classes III to V, pulmonary hemorrhage, myocarditis, major organ vasculitis).

The time-varying predictors included in the visit model were lagged by 1 visit. Model selection was based on fitting a regression model with all available predictors, and subsequently retaining predictors with P-values < 0.05. Analysis used the "coxph" function [48] in R version 3.1.0. Table 2.1 presents the model summary.

Characteristic	**Time-Varying**	**Hazard Ratio**	**95% Hazard Ratio Confidence Limits**	**P-Value**
Disease Activity	Yes	1.02	(1.01, 1.02)	< 0.0001
Prednisone (mg)	Yes	1.02	(1.01, 1.02)	< 0.0001
Composite Score for Significant Immuno-Suppression	Yes	1.07	(1.03, 1.11)	0.001
Ethnicity	No			
Asian Vs. Caucasian	-	1.10	(1.05, 1.15)	< 0.0001
Black Vs. Caucasian	-	1.20	(1.13, 1.26)	< 0.0001
Other Vs. Caucasian	-	1.08	(1.02, 1.14)	0.006
Age at Diagnosis (Years)	No	1.02	(1.01, 1.03)	< 0.0001
Major Organ Involvement Version 1	No	1.07	(1.03, 1.11)	0.001

Table 2.1: The visit process model for the cSLE study.

The model confirmed that visit intensity was positively associated with disease activity (hazard ratio $= 1.02$, 95% confidence interval: 1.01-1.02). As a result, any regression analyses on disease activity should

incorporate the visit process to account for this association; see [7] for an application of inverse-intensity weighted generalized estimating equations to this data.

The R code for modelling the visit process using the Andersen-Gill formulation and estimating the inverse-intensity weights are provided in Appendix A.

2.5 Discussion

This paper proposes novel visual measures for summarizing the extent of visit irregularity by dividing the time frame of interest into bins and counting the number of individuals with 0, 1, and >1 visits per bin. For the TARGet Kids! study, the mean proportions of individuals with 0 visits per bin were above 0.67 while the mean proportions of individuals with >1 visit per bin were below 0.03. This suggested repeated measures data subject to missingness, and thus reasons for why visits are missing should be explored. If investigators deem missingness to be non-informative, the desired longitudinal outcome can be analyzed using appropriate missing data techniques such as multiple imputation. For the cSLE study, visits were recommended to occur at least once every 6 months. For bin widths of 6 months, the mean proportion of individuals with >1 visit per bin was 0.39 and the mean proportion of individuals with 1 visit per bin was 0.39. The mean proportion of individuals with 1 visit per bin was maximized at bin widths of 3.52 months with a value of 0.48 (the mean proportion of individuals with >1 visit per bin was 0.15 at bin widths of 3.52 months). Semi-parametric regression analyses on visit intensity showed that higher disease activity was associated with more frequent visits, and therefore regression analyses on the outcome should account for the visit process.

Irregular longitudinal data is often mishandled in practice. For example, researchers who know repeated measures ANOVA cannot handle irregular data assume they cannot use the data at all, or can use data from scheduled visits only. The latter approach can protect from bias when the visit process is VNAR; however, it is inefficient when the visit process is VCAR or VAR as outcome information is discarded. Other researchers may be aware that certain methods for longitudinal data (e.g. generalized estimating equations, mixed models) will run on unbalanced visits but falsely assume that the results will be unbiased, so they neglect the visit process and risk biased results. In the cSLE study for example, this would result in bias because individuals visited more frequently when their disease status worsened, and thus an unadjusted GEE analysis risks overestimating the burden of disease.

Visit irregularity and missing data are related concepts; however, the timings of visits are rarely scrutinized [41] whereas exploring missing data is recommended practice (e.g. STROBE, CONSORT 2010) [37] [38]. For example, the STROBE guideline encourages the reporting missing data by "indicating the number of participants with missing data for each variable of interest" [37]. Furthermore, identifying predictors of missingness is also generally recommended, see [44] for an example of how this can be done. Similar to missing data techniques, our measures of irregularity count the number of individuals with 0, 1, and >1 visits in each bin. Fitting a recurrent event regression model for the visit intensity to distinguish between VCAR and VAR is analogous to using logistic regression to identify predictors of missingness.

Judging the visit process is crucial to modelling the outcome; we have presented this in terms of determining whether the visit process is VAR or VCAR; however, this can also be viewed in terms of ignorability. In missing data analysis, Little and Rubin [33] defined ignorability as not needing to model the missing data mechanism (data is missing at random or missing completely at random) when

performing likelihood inference on the outcome. Farewell *et al* [49] extended the concept of ignorability to irregular longitudinal data and showed that stability is a sufficient condition for ignorability. Stability requires the outcome at the j^{th} visit to be independent of any visit patterns conditional on the observed data up to the j^{th} visit. In the presence of ignorability, parametric analyses can ignore the visit process.

Modelling the outcome trajectory using a mixed effects regression model is biased if the visit process depends on past observed outcomes and the covariance between the repeated measures is not correctly specified [50]. Several strategies can handle informative visit processes more effectively. Two main semi-parametric approaches for incorporating the visit process are: jointly modelling the outcome and visit processes using shared random effects [22] and constructing generalized estimating equations where observations are weighted by the inverse of their visit intensity [20]. Each strategy relies on a set of assumptions concerning the relationship between the visit and outcome process in relation to covariates and prior visits and outcomes [15]. Since each strategy was developed for specific visit scenarios, no modelling strategy can accommodate all possible cases. Thus, careful consideration of the visit process and study design should inform the chosen analytic method.

While our proposed measures of irregularity can help to distinguish between repeated measures and irregular data, the specification of bin widths is not always straightforward. Consulting with a clinician may help in such cases. For example, the left side of the bins for the TARGet Kids! study was fixed at 5% of the gap between successive visits because it was understood that well child visits cannot be billed if they occur too early and vaccinations are not administered before a child is a certain age. We have also illustrated that varying bin widths can shed light on the visit process.

With missing data, the proportions of missing values provide an easily interpreted score of how severe the problem is. It would be ideal to have a single number that can be used to indicate the extent of irregularity. We are currently investigating the area under the curve (AUC) obtained from plotting the mean proportions of individuals with 0 visits per bin against the mean proportions of individuals with >1 visit per bin. The AUC is a single number that can be used to describe the extent of irregularity where larger values of the AUC would signify increasing irregularity.

2.6 Conclusions

Describing the extent of irregularity is an important step in determining the correct analytic approach to modelling the outcome. Choosing to ignore irregularity and simply use a mixed effects model leads to bias when the observed history (e.g. past outcomes and visits etc.) is predictive of future visit intensity. Exploring visit irregularity is as important as exploring missing data, and our measures of the extent of irregularity can assist in selecting the appropriate methodology for handling the longitudinal outcome.

Chapter 3

Quantifying the Extent of Visit Irregularity in Longitudinal Data

This manuscript has been submitted for publication.

Abstract

The timings of visits in observational longitudinal data may depend on the study outcome, and this can result in bias if ignored. Assessing the extent of visit irregularity is important because it can help determine whether visits can be treated as repeated measures or as irregular data. We propose plotting the mean proportions of individuals with 0 visits per bin against the mean proportions of individuals with >1 visit per bin as bin width is varied and using the area under the curve (AUC) to assess the extent of irregularity. The AUC is a single score which can be used to quantify the extent of irregularity and assess how closely visits resemble repeated measures. Simulation results confirm that the AUC increases with increasing irregularity while being invariant to sample size and the number of scheduled measurement occasions. A demonstration of the AUC was performed on the TARGet Kids! study which enrolls healthy children aged 0-5 years with the aim of investigating the relationship between early life exposures and later health problems. The quality of statistical analyses can be improved by using the AUC as a guide to select the appropriate analytic outcome approach and minimize the potential for biased results.

3.1 Introduction

In observational longitudinal studies, the timings of visits may be correlated with the study outcome over time, and this can lead to misleading conclusions about the outcome trajectory if ignored [20]. For example, Sun *et al* [25] considered a study on hospital medical costs of chronic heart failure patients in which the analytic objective was to estimate the effects of demographic characteristics such as age on medical costs. It was identified that older patients had lower medical costs at each visit; however, older patients visited more frequently. When the visit process was assumed to be uninformative, the magnitude of the effect of age on medical costs was significantly overestimated compared to an analysis which incorporated the visit process. Irregular data describes the presence of imbalanced visit patterns across individuals (often in the absence of a follow-up schedule). In statistical terms, the timings of visits for each individual are random variables. Our previous paper was the first to discuss how to quantify the extent of irregularity and how to judge whether visits resemble repeated measures [34]. This is an important step in selecting an appropriate statistical approach for the longitudinal outcome [15]. This paper builds on the work from our previous paper and proposes a single estimator for quantifying the extent of irregularity.

Visit irregularity is analogous to the problem of missing data. With missing data, best practice states that rates of missingness should be reported (e.g. STROBE, CONSORT 2010) [37] [38]. Visit irregularity is similar to missing data; however, recent evidence suggests that the extent of irregularity is rarely described and the potential for bias is generally ignored [41].

Assessing the extent of irregularity is important because it can help determine whether it is more appropriate to treat visits as repeated measures or irregular data. There are different methods for modelling longitudinal outcomes when the data are treated as repeated measures (e.g. generalized estimating equations (GEEs) [9], linear mixed effects models [10]), or as irregular data (e.g. inverse-intensity weighted generalized estimating equations [20]). Perfect repeated measures describes the scenario when every individual adheres to a common set of visit times pre-specified by protocol. At the other extreme, the extent of irregularity can vary to the point of no shared visit times across individuals. In practice, there are scenarios between the extremes where there is variability in scheduled visit timings, missed scheduled visits, and unscheduled visits. To determine whether visits should be treated as repeated measures or as irregular data, a more thorough assessment of irregularity is required. This includes an exploration of predictors of visit intensity (e.g. using recurrent event analyses), and the documentation of the extent of irregularity. Our previous paper proposed dividing the timeframe of follow-up into bins, and reporting the proportions of individuals with 0, 1, and >1 visits within each bin [34]. When there is no obvious choice of bin width, we suggested varying bin widths.

It would be ideal to have a single measure that can be used to interpret the extent of irregularity. With missing data, the proportion of observations with missing values provides an easily interpreted measure of how severe the problem is. With irregular data, we propose plotting the mean proportions of individuals with 0 visits per bin against the mean proportions of individuals with >1 visit per bin as bin width is varied and using the area under the curve (AUC) to assess the extent of irregularity. We explore how the AUC behaves using simulations and apply the AUC to the TARGet Kids! study [8].

3.2 The AUC as a Measure of Irregularity

We calculate the AUC by constructing bins across the duration of the study. The widths of bins are varied and the mean proportions of individuals with 0 and >1 visits per bin are calculated at each width. As bins become wider, the mean proportions of individuals with >1 visit per bin increase and the mean proportions of individuals with 0 visits per bin decrease. The mean proportions of individuals with 0 visits per bin are plotted (y-axis) against the mean proportions of individuals with >1 visit per bin (x-axis) and the area under the curve (AUC) is estimated. Figure 3.1 illustrates two mean proportions of individuals with 0 vs. >1 visits per bin curves (and resulting AUCs) derived from two simulated datasets with 1000 observations each.

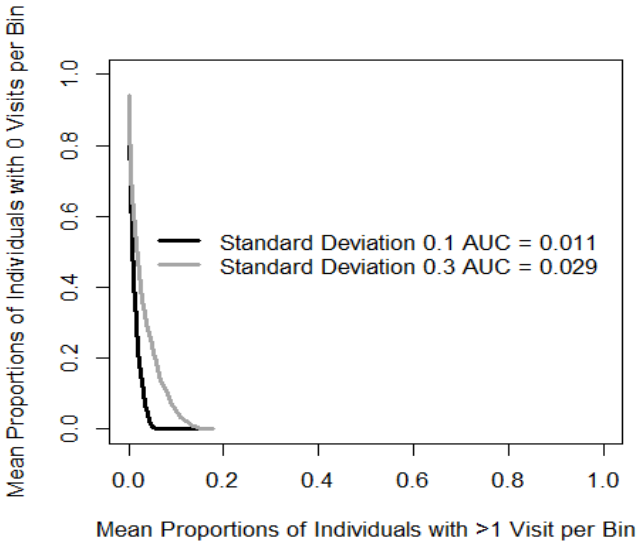

Figure 3.1: The mean proportions of individuals with 0 vs. >1 visits per bin curve and AUCs for repeated measures data for two values of the standard deviation of the timings of scheduled visits.

Both datasets had five pre-specified visit times (2, 4, 6, 8, 10 months), no missing scheduled visits, and unscheduled visits from a homogeneous Poisson process with a rate of 0.1. Both datasets had normally distributed scheduled visit timings centred at each scheduled measurement occasion. The first dataset had a standard deviation of 0.1, and the second dataset had a standard deviation of 0.3. The first dataset had a smaller AUC compared to the second dataset (0.011 vs. 0.029). With repeated measures data, the mean proportions of individuals with >1 visit per bin are low and thus the AUC will be near 0. As the standard deviation of scheduled visit timings increases, visits become more irregular and the AUC will increase. We propose using the AUC to describe the extent of irregularity.

To be an effective measure of irregularity, the AUC estimator should increase with increasing irregularity, be straightforward to interpret, and be invariant to sample size and follow-up length. When there are pre-specified visit times, the AUC should provide an insight into whether visits resemble repeated measures.

When study protocols do not pre-specify visit times, we construct bins using the hazard function of the visit process. Let $\lambda(t)$ be the visit hazard at time t and let $\Lambda(t) = \int_0^t \lambda(u)du$ be the corresponding

cumulative hazard. Let the study duration be τ and let k denote the number of bins. Bin cut-points c_j are specified such that the area under the hazard curve within each bin $(c_{j-1}, c_j]$ is $\Lambda(\tau)/k$. The j^{th} cut-point is obtained by solving the system of equations $\int_{c_{j-1}}^{c_j} \lambda(u)du = \Lambda(\tau)/k$ or equivalently $\Lambda(c_j) - \Lambda(c_{j-1}) = \Lambda(\tau)/k$. We vary bin widths by varying k.

When visit times are pre-specified by protocol, the j^{th} bin is constructed around the j^{th} pre-specified visit time T_j. Let L_j and R_j represent the corresponding left and right cut-points of the j^{th} bin (Figure 3.2).

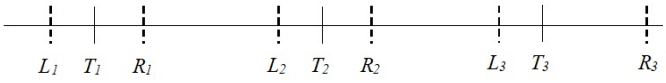

Figure 3.2: Bins around pre-specified visit times.

For the remainder of this paper, L_j is specified as the percentage of the gap between T_{j-1} and T_j, and R_j is specified as the percentage of the gap between T_j and T_{j+1}. The bins are at their widest when $R_{j-1} = L_j$.

3.2.1 Properties of the AUC

If the underlying visit process is a Poisson process with rate function $\lambda(t)$, the AUC will be 0.25 when the bins are specified using the hazard function of visits. To see this, let $N(c_{j-1}, c_j]$ represent the number of visits in bin $(c_{j-1}, c_j]$ for an individual, then the expressions for the probability of observing 0 and >1 visits become:

$$\begin{aligned} P(N(c_{j-1}, c_j] = 0) &= \exp\left(-\int_{c_{j-1}}^{c_j} \lambda(u)du\right) = \exp\left(-\frac{\Lambda(\tau)}{k}\right) \\ P(N(c_{j-1}, c_j] > 1) &= 1 - \exp\left(-\int_{c_{j-1}}^{c_j} \lambda(u)du\right) - \left(\int_{c_{j-1}}^{c_j} \lambda(u)du\right) \exp\left(-\int_{c_{j-1}}^{c_j} \lambda(u)du\right) \\ &= 1 - \exp\left(-\frac{\Lambda(\tau)}{k}\right) - \left(\frac{\Lambda(\tau)}{k}\right) \exp\left(-\frac{\Lambda(\tau)}{k}\right) \end{aligned} \tag{3.1}$$

The mean proportions of 0 and >1 visits per bin are equivalent to the respective probability expressions above. Writing $x = P(N(c_{j-1}, c_j] = 0)$, then $P(N(c_{j-1}, c_j] > 1) = 1 - x + x\log(x)$ and the AUC is:

$$\int_0^1 1 - x + x\log(x)dx = 0.25$$

Figure 3.3 illustrates different repeated measures scenarios and how the AUC relates to missingness, variability in scheduled visit timings, and unscheduled visits.

When data are perfect repeated measures (no variability in scheduled visit timings, no missing data, and no unscheduled visits), the AUC is 0 because the mean proportions of individuals with 0 vs. >1 visits per bin curve is a point at (0, 0) (Figure 3.3i). If variability in scheduled visit timings is introduced, the AUC is still 0 because the mean proportions of individuals with >1 visit per bin are 0 for all bin

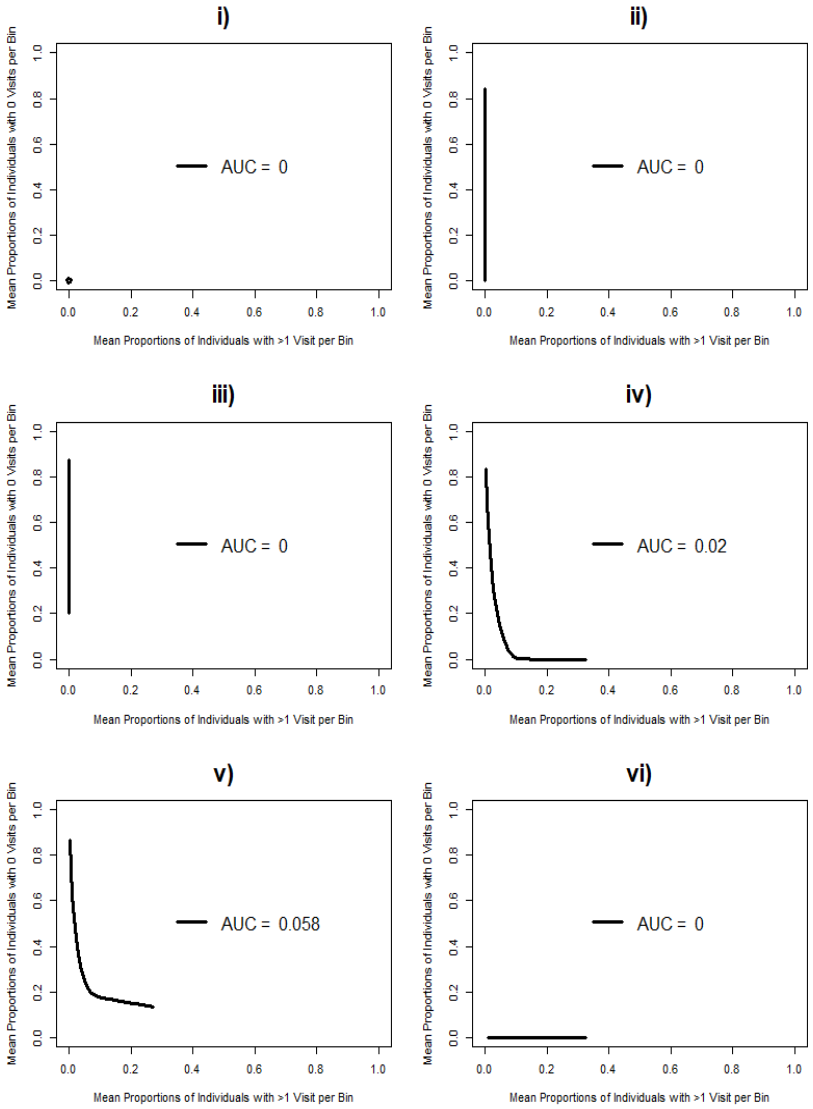

Figure 3.3: The mean proportions of individuals with 0 vs. >1 visits per bin and AUCs for repeated measures with and without missingness, variability in scheduled visit timings, and unscheduled visits.

widths (Figure 3.3ii) (unless the extent of variability is so large such that some scheduled visits end up in other bins). If there is variability in scheduled visit timings and missing scheduled visits but there are no unscheduled visits, the AUC is 0 because the mean proportions of individuals with >1 visit per bin remain at 0 for all bin widths (Figure 3.3iii). If unscheduled visits are added to the scenario in Figure 3.3ii (variability in scheduled visit timings and no missingness), the AUC increases because there will be mean proportions of individuals with >1 visit per bin above 0 (Figure 3.3iv). If missingness is also introduced, the AUC further increases because there will be increased mean proportions of individuals with 0 visits per bin (Figure 3.3v). The AUC can be 0 in the presence of unscheduled visits. This occurs when there is no missingness and no variability in scheduled visit timings because the mean proportions of individuals with 0 visits per bin are 0 for all bin widths (Figure 3.3vi). In this scenario, we recommend exploring unscheduled visits separately from the scheduled visits. In general, we recommend reporting the AUC together with the mean proportions of individuals with 0 vs. >1 visits per bin plot.

The AUC can assume values between 0 and 0.5 (the latter is the AUC of a straight line with a y-intercept and x-intercept of 1). There are two cases in which the AUC asymptotically approaches 0.5. The first case is when some individuals have 0 visits, and other individuals have a visit followed instantly by another visit. The second case is when some individuals have 0 visits, and other individuals have a visit rate of λ which approaches infinity.

3.2.2 Estimating the AUC in the Absence of Missingness

With repeated measures data, the AUC increases as the level of missingness increases. This is problematic because: 1) different combinations of missingness and unscheduled visit rates can yield the same AUC, and 2) the AUC is intended to provide insight into the extent of irregularity, not missingness. To eliminate the influence of missingness on the AUC and allow judgement solely on the extent of irregularity, we derive the AUC had there been no missing data using likelihood-based estimation (we denote this AUC as AUC0). Let k denote the number of scheduled measurement occasions. To formulate the likelihood function for the i^{th} individual ($i = 1$ to n) in the j^{th} bin ($j = 1$ to k) ($\mathcal{L}_{ij}$), we assume three independent processes: 1) the timings of scheduled visits around a pre-specified visit time, 2) whether scheduled visits are missing, and 3) unscheduled visits. Let the distribution of the scheduled visit time for bin j have density function $f_j(t;\mu_j)$ (where μ_j is a vector of parameters) with support in the widest possible bin (i.e. $f_j(t;\mu_j)$ is 0 outside of the widest possible cut-points for bin j) and cumulative distribution function $F_j(t;\mu_j)$. Let whether a scheduled visit was missing or not be a Bernoulli random variable with probability of missingness being $\pi_j(\nu_j;B_{ij})$ (where ν_j is a vector of regression parameters, and B_{ij} is a vector of covariates). Let the rate function of unscheduled visits for subject i in bin j be $\lambda_{ij}(t;\alpha_j)$ (where α_j is a vector of parameters) which can incorporate certain assumptions (e.g. $\lambda_{ij}(t;\alpha_j) = \lambda_j(t;\alpha_j)$). The rate function is specified parametrically to allow for an informative unscheduled visit process (e.g. model $\lambda_{ij}(t;\alpha_j)$ as a function of past covariates, outcomes etc.). Let n_{ij} denote the number of visits for the i^{th} individual in the j^{th} bin. Let the observed visit times for the i^{th} individual in the j^{th} bin be $t_{ij1},\ldots,t_{ijn_{ij}}$.

To derive $\mathcal{L}_{ij}$, we consider 2 cases: 1) a scheduled visit was missing, and 2) a scheduled visit was observed. Suppose a scheduled visit was missing ($I()$ is the indicator function), then the likelihood contribution given n_{ij} ($\mathcal{L}_{ij}^{\text{missing}}$) is:

If $n_{ij} = 0$ then:

$$\mathcal{L}_{ij}^{\text{missing}} = \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)$$

If $n_{ij} = 1$ then:

$$\mathcal{L}_{ij}^{\text{missing}} = \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\lambda_{ij}(t_{ij1};\alpha_j)$$

In general:

$$\begin{aligned}\mathcal{L}_{ij}^{\text{missing}} &= I(n_{ij}=0)\exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) + I(n_{ij}>0)\exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\prod_{r=1}^{n_{ij}}\lambda_{ij}(t_{ijr};\alpha_j)\\ &= \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\left[I(n_{ij}=0) + I(n_{ij}>0)\prod_{r=1}^{n_{ij}}\lambda_{ij}(t_{ijr};\alpha_j)\right]\end{aligned} \tag{3.2}$$

If a scheduled visit occurs, then the likelihood contribution given n_{ij} ($\mathcal{L}_{ij}^{\text{scheduled visit}}$) is:

If $n_{ij} = 2$ then:

$$\begin{aligned}\mathcal{L}_{ij}^{\text{scheduled visit}} &= \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\left[f_j(t_{ij1};\mu_j)\lambda_{ij}(t_{ij2};\alpha_j) + f_j(t_{ij2};\mu_j)\lambda_{ij}(t_{ij1};\alpha_j)\right]\\ &= \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\prod_{r=1}^{2}\left[\lambda_{ij}(t_{ijr};\alpha_j)\left(\sum_{q=1}^{2}\frac{f_j(t_{ijq};\mu_j)}{\lambda_{ij}(t_{ijq};\alpha_j)}\right)\right]\end{aligned} \tag{3.3}$$

In general:

$$\mathcal{L}_{ij}^{\text{scheduled visit}} = \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\prod_{r=1}^{n_{ij}}\left[\lambda_{ij}(t_{ijr};\alpha_j)\left(\sum_{q=1}^{n_{ij}}\frac{f_j(t_{ijq};\mu_j)}{\lambda_{ij}(t_{ijq};\alpha_j)}\right)\right]$$

The general expressions are combined to derive $\mathcal{L}_{ij}$:

$$\begin{aligned}\mathcal{L}_{ij} &= \pi_j(\nu_j;B_{ij})\mathcal{L}_{ij}^{\text{missing}} + (1-\pi_j(\nu_j;B_{ij}))\mathcal{L}_{ij}^{\text{scheduled visit}}\\ &= \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right)\left[\pi_j(\nu_j;B_{ij})I(n_{ij}=0) + I(n_{ij}>0)\left(\pi_j(\nu_j;B_{ij})\prod_{r=1}^{n_{ij}}[\lambda_{ij}(t_{ijr};\alpha_j)]\right.\right.\\ &\left.\left. + (1-\pi_j(\nu_j;B_{ij}))\prod_{r=1}^{n_{ij}}\left[\lambda_{ij}(t_{ijr};\alpha_j)\left(\sum_{q=1}^{n_{ij}}\frac{f_j(t_{ijq};\mu_j)}{\lambda_{ij}(t_{ijq};\alpha_j)}\right)\right]\right)\right]\end{aligned} \tag{3.4}$$

The likelihood function can be maximized to obtain estimates of ν_j, μ_j, and α_j. The expressions for the probabilities of 0 visits per bin $P_j(0)$ and 1 visit per bin $P_j(1)$ are:

$$
\begin{aligned}
P_j(0) &= \pi_j(\nu_j; B_{ij}) P(\text{no unscheduled visits in bin } j \mid \text{missing scheduled visit}) \\
&+(1-\pi_j(\nu_j; B_{ij})) P(\text{no unscheduled visits in bin } j \text{ and scheduled visit outside bin } j \mid \text{scheduled visit}) \\
&= \pi_j(\nu_j; B_{ij}) \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) \\
&+(1-\pi_j(\nu_j; B_{ij})) \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) (F_j(L_j;\mu_j) + 1 - F_j(R_j;\mu_j))
\end{aligned}
$$

$$
\begin{aligned}
P_j(1) &= \pi_j(\nu_j; B_{ij}) P(1 \text{ unscheduled visit in bin } j \mid \text{missing scheduled visit}) \\
&+(1-\pi_j(\nu_j; B_{ij})) P(1 \text{ unscheduled visit in bin } j \text{ and scheduled visit outside bin } j \mid \text{scheduled visit}) \\
&+(1-\pi_j(\nu_j; B_{ij})) P(\text{no unscheduled visits in bin } j \text{ and scheduled visit in bin } j \mid \text{scheduled visit}) \\
&= \pi_j(\nu_j; B_{ij}) \left(\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) \\
&+(1-\pi_j(\nu_j; B_{ij})) \left(\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) (F_j(L_j;\mu_j) + 1 - F_j(R_j;\mu_j)) \\
&+(1-\pi_j(\nu_j; B_{ij})) \exp\left(-\int_{L_j}^{R_j} \lambda_{ij}(u;\alpha_j)du\right) (F_j(R_j;\mu_j) - F_j(L_j;\mu_j))
\end{aligned}
$$

Given the probability of >1 visit per bin $P_j(>1) = 1 - P_j(0) - P_j(1)$, the mean proportions of individuals with 0, 1, and >1 visits per bin can be obtained by averaging $P_j(0), P_j(1)$ and $P_j(>1)$ across bins respectively. To obtain an AUC assuming no missingness (AUC0), the expressions for $P_j(0), P_j(1)$ and $P_j(>1)$ are evaluated at $\pi_j(\nu_j; B_{ij}) = 0$ and the corresponding likelihood estimates of μ_j and α_j.

3.3 Simulations

When there are no pre-specified visit times, we assessed whether the AUC is invariant to sample size and study duration, and whether the AUC increases as the variance of the gap times between successive visits increases. In practice, visit rates may not be the same across individuals so we evaluated whether the AUC increases as the degree of visit heterogeneity increases.

When there are pre-specified visit times, we investigated whether the AUC is invariant to sample size, the level of missingness, and number of scheduled measurement occasions, and whether the AUC increases as the standard deviation of scheduled visit timings and rate of unscheduled visits increase. We also explored scenarios where unscheduled visits impact whether scheduled visits are missing, and when the rate of unscheduled visits increases due to an unobserved process.

3.3.1 Simulation Methods

No Pre-Specified Visit Times

To evaluate whether the AUC increases as the variance of gap times increases, we simulated 1000 datasets with each dataset having n = 30, 100 and a study duration of $\tau = 15$, 30 months. The parameters of the gap time distribution were specified such that we could vary the variance while simultaneously holding the mean constant at 4 months. If the gap times follow a Log-normal(μ, σ^2) distribution, then the mean is $\exp\left(\mu + \frac{\sigma^2}{2}\right)$ and the variance is $\exp(2\mu + \sigma^2) \times (\exp(\sigma^2) - 1)$. The Log-normal parameters were specified to be $\sigma^2 = 0.1$, 0.2, 0.3, 0.4, 0.5 and $\mu = \log 4 - \frac{\sigma^2}{2}$. If the gap times folllow a Gamma(α, β) distribution, then the mean is $\frac{\alpha}{\beta}$ and the variance is $\frac{\alpha}{\beta^2}$. The Gamma parameters were specified as $\alpha = 0.8v$ and $\beta = \frac{\alpha}{4}$ for $v = 1$, 2, 3, 4, 5.

To investigate the effect of visit heterogeneity on the AUC, we simulated 1000 datasets with each dataset having $n = 30$, 100 and the study duration was held constant at $\tau = 30$ months. Visits were simulated from two homogenous Poisson processes with half of the individuals being randomly selected to have rate $\lambda_1 = 0.1$, 0.5, 0.9 and the other half rate $\lambda_2 = 1.0$, 1.4, 1.8 respectively. The distance between the two rates was varied such that $(\lambda_1, \lambda_2) = (0.9, 1.0)$, $(0.5, 1.4)$, $(0.1, 1.8)$.

Pre-Specified Visit Times

For pre-specified visit times, we compared the following AUCs: 1) the AUC based on the observed mean proportions of individuals with 0 vs. >1 visits per bin (AUC_{OBS}), and 2) the likelihood-based AUC (AUC_{MLE}). Likelihood-based AUCs assuming no missingness (AUC0) were also estimated. We simulated 1000 datasets with each dataset having three (2, 4, 6 months) or five (2, 4, 6, 8, 10 months) scheduled measurement occasions and sample size $n = 30$, 100. Each dataset was simulated using three independent processes: 1) timings of scheduled visits, 2) missing scheduled visits, and 3) unscheduled visits. The timings of scheduled visits were simulated as normal random variables centred at each scheduled measurement occasion with a standard deviation of $\sigma = 0.1$, 0.3. Missing scheduled visits within each bin were simulated as Bernoulli random variables with the same probability of missing a scheduled visit $\pi = 0.1$, 0.2, 0.3 for all bins. Unscheduled visits were simulated from a homogenous Poisson process with rate $\lambda = 0.1$, 0.15, 0.2, 0.25, 0.3 (and bins were assumed to be independent). To include as much visit information as possible, the likelihood function was maximized using the widest bins.

We considered two departures from the above set up. As independence among the missingness, scheduled visit timings, and unscheduled visit processes is unlikely in practice, the first departure we considered was when the probability of a scheduled visit being missing increases when an unscheduled visit occurs just before it. To explore this violation of the assumption of independence between the scheduled and unscheduled visit processes, we simulated 1000 datasets with each dataset having three scheduled measurement occasions (2, 4, 6 months) and $n = 100$. The timings of scheduled visits were simulated as normal random variables centred at each scheduled measurement occasion with a standard deviation of $\sigma = 0.05$, 0.2. Unscheduled visits were simulated from a homogenous Poisson process with rate $\lambda = 0.1$, 0.3. The probability of missing a scheduled visit was $\pi + \theta I(\text{at least 1 unscheduled visit in } (T_{j-0.5}, T_j])$ where $0 \leq \pi + \theta \leq 1$ and where $I()$ is the indicator function. The parameter π was specified as 0.1, and the shift parameter θ was specified as $0.1, 0.5$.

We compared the correct AUC_{MLE} (and AUC0) to: 1) a mis-specified AUC_{MLE} (and AUC0) assuming the independence structure in the beginning of this section, and 2) AUC_{OBS}.

The second departure we considered was when an unobserved process triggers an increase in the rate of unscheduled visits. For example, in a study of patients with childhood-onset Systemic Lupus Erythematosus [7], an increase in disease activity was suspected to lead to an increased rate of visits. To evaluate the AUC in this setting, we simulated 1000 datasets with each dataset having three scheduled measurement occasions (2, 4, 6 months) with $n = 100$. The timings of scheduled visits were simulated as normal random variables centred at each scheduled measurement occasion with a standard deviation of $\sigma = 0.1$. The probability of missing a scheduled visit was $\pi = 0.1$. Unscheduled visits were simulated from a homogenous Poisson process with rate $\lambda = 0.1, 0.2$, and increased to $\lambda_{act} = 0.4, 0.6$. The potential activation times were simulated from an independent homogenous Poisson process with rate λ_{act} wherein the probability of activating the increased rate of unscheduled visits at each time was $P_{act} = 0.3, 0.6$, and the probability of the increased rate of unscheduled visits deactivating at each time was $P_{deact} = 0.1, 0.3$.

To assess the consequences of mis-specifying the likelihood function in this scenario, we compared an AUC_{OBS} to a mis-specified AUC_{MLE} assuming an independence structure and the corresponding AUC0.

To evaluate the relationship between the AUC and bias, we define the outcome $Y_i(t)$ for the i^{th} individual at time t as follows:

$$Y_i(t) = u_i + \epsilon_i(t) \tag{3.5}$$

where the random intercept u_i and the random error at time t $\epsilon_i(t)$ follow normal distributions with mean 0 and variance 0.5 and are independent of each other (the random errors $\epsilon_i(t)$ are independent and identically distributed (i.i.d)), and thus the outcome $Y_i(t)$ follows a normal distribution with mean 0 and variance 1.

We simulated 1000 datasets with each dataset having two (5, 10 months) or four (2.5, 5, 7.5, 10 months) scheduled measurement occasions and $n = 100$. Two processes contributed to variability in visit times among individuals: 1) timings of scheduled visits, and 2) unscheduled visits. It was assumed that there were no missing scheduled visits. The timings of scheduled visits were simulated as normal random variables centred at each scheduled measurement occasion with a standard deviation of $\sigma = 0.1$. Let U_{ik} denote the k^{th} unscheduled visit time and let S_{ij} denote the j^{th} scheduled visit time for subject i. To simulate U_{ik}, the gap times between successive unscheduled visits $(U_{ik} - U_{ik-1})$ were simulated as exponential random variables with rate $\lambda_0 \exp(\gamma Y_i(U_{ik-1}))$, where the baseline rate $\lambda_0 = 0.1, 0.2, 0.3, 0.4$, and the parameter $\gamma = 0, 0.5, 1$. However, if a scheduled visit (S_{ij}) occurred between the two successive unscheduled visits, then U_{ik} would reset and the gap time between the scheduled visit and the unscheduled visit $(U_{ik} - S_{ij})$ was simulated as an exponential random variable with rate $\lambda_0 \exp(\gamma Y_i(S_{ij}))$. The unscheduled visit times (U_{ik}) could reset more than once.

We assessed the impact of the level of informativeness of the unscheduled visit process (which increases as γ increases) on the AUC_{OBS} and the level of bias (i.e. the mean of the outcomes).

3.3.2 Simulation Results

The results of all of the performed simulations are summarized in tables in Appendix B.

No Pre-Specified Visit Times

The AUC increased as the extent of irregularity increased (this was represented by larger values of the gap time variance) (see Figures 3.4 and 3.5). For each value of τ, comparing $n = 30$ and $n = 100$ yielded similar mean AUCs. For larger gap time variances, the mean AUCs were larger for $\tau = 30$ months. For smaller gap time variances, the mean AUCs were larger for $\tau = 15$ months. The results evaluating the impact of visit heterogeneity on the AUC are presented in Table 3.1. The mean AUCs increased as the distance between λ_1 and λ_2 increased. Within each specification of (λ_1, λ_2), the mean AUCs were similar as the sample size increased.

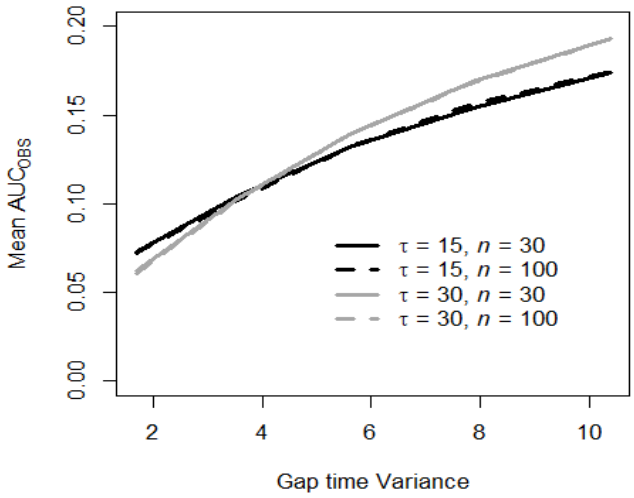

Figure 3.4: The mean observed AUCs (AUC_{OBS}) for Log-normal gap times across the study duration (τ), sample size (n), and gap time variance.

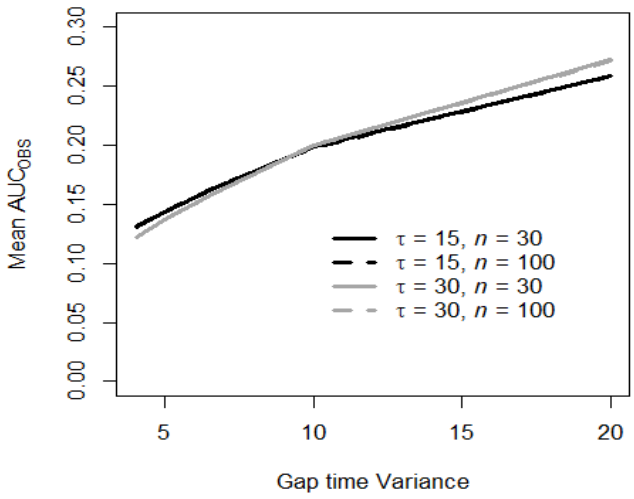

Figure 3.5: The mean observed AUCs (AUC_{OBS}) for Gamma gap times across the study duration (τ), sample size (n), and gap time variance.

(λ_1, λ_2)	n	**Mean AUC_{OBS}**	**Standard Error**
(0.9, 1.0)	30	0.250	0.007
	100	0.250	0.004
(0.5, 1.4)	30	0.277	0.008
	100	0.278	0.005
(0.1, 1.8)	30	0.343	0.008
	100	0.345	0.004

Table 3.1: The mean observed AUCs (AUC_{OBS}) across visit rates (λ_1, λ_2), and sample size (n).

Pre-Specified Visit Times

The mean AUCs increased with increasing irregularity (this was represented by larger values of σ and λ), but were invariant to sample size and the number of scheduled measurement occasions (Figure 3.6). While the mean AUCs increased with increasing missingness, the mean AUC0 were invariant to the level of missingness (Figure 3.7).

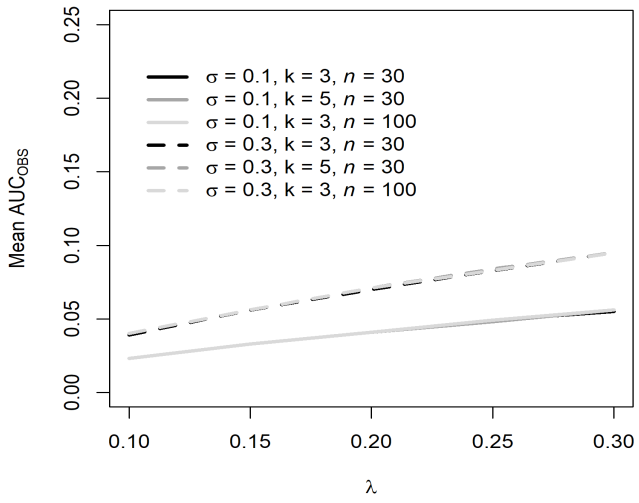

Figure 3.6: The mean observed AUCs (AUC_{OBS}) across the rate of unscheduled visits (λ), sample size (n), number of scheduled measurement occasions (k), and the standard deviation of scheduled visit timings (σ).

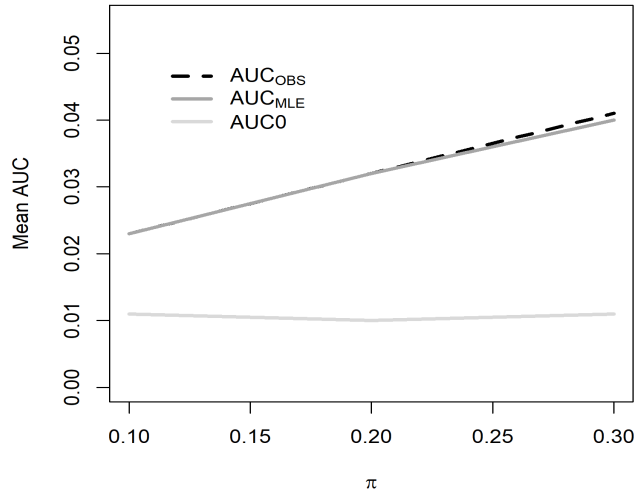

Figure 3.7: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) across the level of missingness (π).

Figures 3.8 and 3.9 display the results for the case where unscheduled visits caused an increase in the probability of scheduled visits being missing. The mean AUCs increased with increasing irregularity (this was represented by larger values of σ, and smaller values of θ because more individuals will have >1 visit in a bin as θ decreases) (Figure 3.8). For $\sigma = 0.05$, the mean AUC_{OBS} and correct mean AUC_{MLE}

were similar. For $\sigma = 0.2$, the correct mean AUC_{MLE} diverged from the mean AUC_{OBS} as θ increased. The correct mean AUC0 were invariant to θ while the mis-specified mean AUC0 decreased as θ increased for $\sigma = 0.2$ (Figure 3.9).

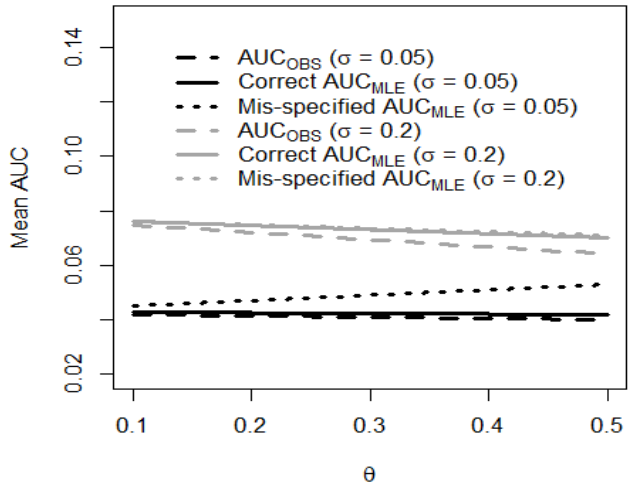

Figure 3.8: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE}) across the increase in the probability of missingness due to unscheduled visits (θ), and the standard deviation of scheduled visit timings (σ).

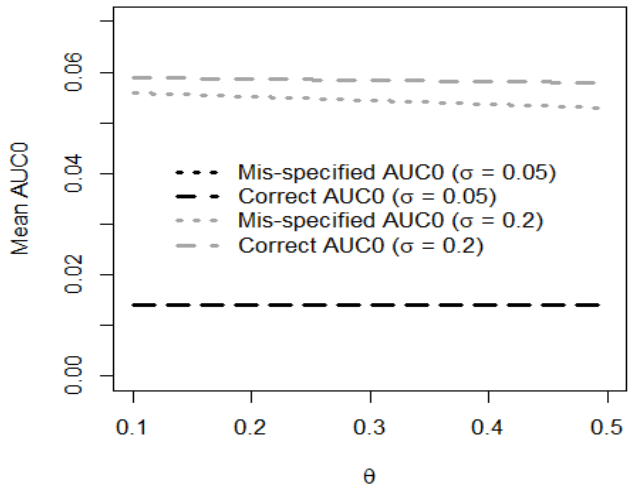

Figure 3.9: The correct and mis-specified likelihood-based mean AUCs assuming no missingness (AUC0) across the increase in the probability of missingness due to unscheduled visits (θ), and the standard deviation of scheduled visit timings (σ).

Figure 3.10 presents the results when the rate of unscheduled visits increased due to an unobserved process. The mean AUCs increased as irregularity increased (this was represented by larger values of λ and λ_{act}). The mean AUCs increased as P_{act} increased because the higher rate of unscheduled visits was more likely to be triggered earlier in time (and the mean AUCs decreased as P_{deact} increased). The mean AUC_{OBS} and the mis-specified mean AUC_{MLE} were similar because the mis-specified likelihood-based estimate of the single rate of unscheduled visits was a weighted average of λ and λ_{act} (see Table B.8 in Appendix B).

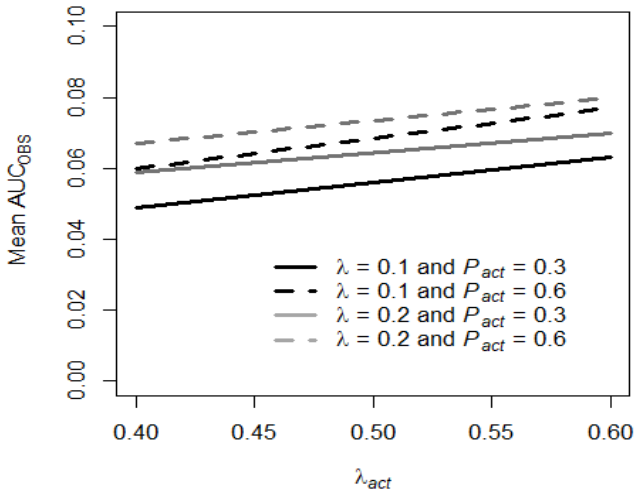

Figure 3.10: The mean observed AUCs (AUC_{OBS}) across the initial rate of unscheduled visits (λ), the increased rate of unscheduled visits (λ_{act}), and the probability of activating the increased rate of unscheduled visits (P_{act}).

Regarding the relationship between the AUC and bias, when the unscheduled visit process was informative (i.e. $\gamma \neq 0$), the mean AUC_{OBS} increased as the mean bias increased, while for an uninformative unscheduled visit process (i.e. $\gamma = 0$), the mean bias was approximately 0 as the mean AUC_{OBS} increased (Figure 3.11).

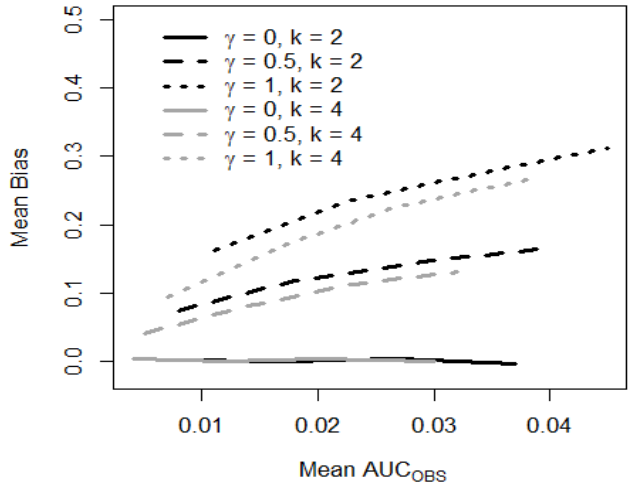

Figure 3.11: The relationship between mean bias and the mean observed AUCs (AUC_{OBS}) across the number of scheduled measurement occasions (k), and the increase in the log rate of unscheduled visits for a standard deviation increase in the outcome (γ) for sample size (n) 100.

3.4 The Magnitude of the AUC in Relation to Bias

To assess whether data can be viewed as repeated measures, the AUC is compared to 0; however, it can be difficult to decide how large of an AUC is tolerable to treat the data as repeated measures. The objective of this section is to help researchers determine acceptable thresholds for the AUC within a repeated measures setting for a given level of bias. The framework of this section is similar to the simulation study which evaluated the relationship between the AUC and bias in the pre-specified visit times case (Section 3.3.1). This section explores a wider range of parameter values corresponding to scheduled visit timings and unscheduled visits, and thus explores a wider range of AUC values. The outcome definition for the i^{th} individual at time t is:

$$Y_i(t) = u_i + \epsilon_i(t)$$

where the random intercept u_i and the random error at time t $\epsilon_i(t)$ follow normal distributions with mean 0 and variance 0.5 and are independent of each other (the random errors $\epsilon_i(t)$ are independent and identically distributed (i.i.d)). The outcome $Y_i(t)$ follows a normal distribution with mean 0 and variance 1.

3.4.1 Simulation Methods

Each simulated dataset had either two $(5, 10$ months) or four $(2.5, 5, 7.5, 10$ months) scheduled measurement occasions and $n = 100$. For each simulated dataset, visits were generated using two processes: 1) timings of scheduled visits (there were no missing scheduled visits), and 2) unscheduled visits. The timings of scheduled visits were simulated as normal random variables centred at each

scheduled measurement occasion with a standard deviation of $\sigma = 0.1, 0.3, 0.6$. The gap times between successive unscheduled visits $(U_{ik} - U_{ik-1})$ for subject i were simulated as exponential random variables with rate $\lambda_0^* \exp(\gamma Y_i(U_{ik-1}))$, where the baseline rate $\lambda_0^* = \lambda_0 v_i$ incorporates a frailty term v_i which was simulated from a Gamma$(1, 1)$ distribution, and was then multiplied by $\lambda_0 = 0.05, 0.10, 0.25, 0.40, 0.60, 1.10, 1.30, 2.00$. The level of informativeness of the unscheduled visit process was varied ($\gamma = 0, 0.5, 1$). Similar to the previous simulation setup (Section 3.3.1), U_{ik} would reset if a scheduled visit (S_{ij}) occurred between two successive unscheduled visits, and the gap time between the scheduled visit and the unscheduled visit $(U_{ik} - S_{ij})$ was simulated as an exponential random variable with rate $\lambda_0^* \exp(\gamma Y_i(S_{ij}))$. Table 3.2 displays the possible values of each parameter.

Visit Process Components	**Parameters**	**Parameter Values**
Scheduled Visit Timings	σ	$0.1, 0.3, 0.6$
	k	$2, 4$
Unscheduled Visits	λ_0	$0.05, 0.10, 0.25, 0.40, 0.60, 1.10, 1.30, 2.00$
	γ	$0.0, 0.5, 1.0$

Table 3.2: The parameter values corresponding to the timings of scheduled visits and unscheduled visits for the simulated repeated measures datasets.

We simulated 1000 datasets for each combination of the parameter values in Table 3.2. We assessed the relationship between the mean AUC_{OBS} and the level of bias (i.e. the mean of the outcomes) for different values of σ, k, λ_0 and γ.

3.4.2 Simulation Results

When the unscheduled visit process was informative (i.e. $\gamma \neq 0$), the largest magnitude of the mean bias was 0.43, and the range of the mean AUC_{OBS} was between 0.002 and 0.26. Figure 3.12 visualizes the relationship between the mean AUC_{OBS} (on the log scale) and mean bias across the number of scheduled measurement occasions (k) for an informative unscheduled visit process (i.e. $\gamma = 0.5, 1$).

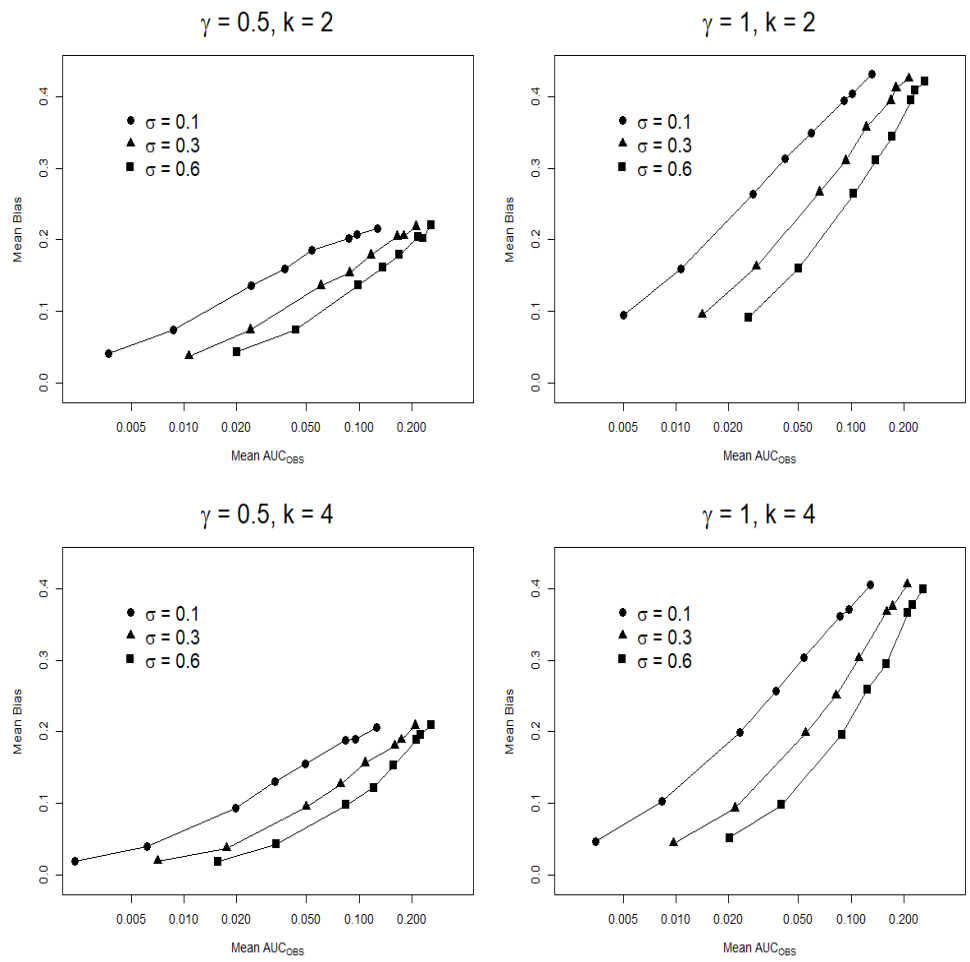

Figure 3.12: The mean bias plotted against the mean observed AUCs (AUC_{OBS}) on the log scale across the number of scheduled measurement occasions (k), and the standard deviation of scheduled visit timings (σ) for an informative unscheduled visit process ($\gamma = 0.5, 1$).

The scatter plots indicated a positive relationship between the mean AUC_{OBS} and mean bias when the unscheduled visit process was informative. For similar values of the mean AUC_{OBS}, the mean bias values were lower when there were four scheduled measurement occasions compared to two scheduled measurement occasions. Furthermore, the mean bias values were larger for a given mean AUC_{OBS} when it was λ_0 and not the variability in scheduled visit timings (σ) that was the greatest contributor to the extent of irregularity. The relationship between the mean bias and the mean AUC_{OBS} on the log scale was nearly linear, but there were slower initial increases in mean bias for a unit increase in log mean AUC_{OBS}, and then there were larger increases in mean bias with an eventual levelling off. The larger increases in mean bias happened when the mean AUC_{OBS} was larger than 0.005, and the levelling off happened when the mean AUC_{OBS} was larger than 0.08 (the mean bias increased by at most 0.17 as the mean AUC_{OBS} increased from 0.08 to 0.26). In terms of thresholds for the AUC, values below 0.005 could be considered small, and values above 0.08 could be considered large in this context. The mean bias was always less than 0.1 when the AUC was below 0.005. Overall, as the level of informativeness increased, the mean bias increased; however, there were no distinguishing mean AUC_{OBS} values which

could be clearly tied to a certain level of informativeness, and hence a level of bias. Plots of the mean proportions of individuals with 0 vs. >1 visits per bin for randomly selected simulated datasets are presented in Figures B.1 and B.2 in Appendix B.

The simulation results reaffirmed the complex relationship between bias and the AUC. The first step in judging the AUC in a repeated measures setting should be to evaluate the level of informativeness of the unscheduled visit process. Following this, the figures in this section can be referred to in order to find the AUC value at which the level of bias becomes unacceptable. We have demonstrated that unscheduled visits and variability in scheduled visit timings can have different impacts on bias. This points towards a strength of the likelihood-based AUC estimator (AUC_{MLE}) as these two visit components are considered separately, and the resulting parameter estimates can provide a summary of each component's contribution to the extent of irregularity. This can provide further insight into the potential for bias and can help determine whether the data can be treated as repeated measures (e.g. if the estimate of the rate of unscheduled visits is small relative to the standard deviation of scheduled visit timings, the observed irregularity could potentially be ignored, but not vice versa).

3.5 Application of the AUC

3.5.1 TARGet Kids! Study

The TARGet Kids! study follows a cohort of children to investigate the impacts of early life exposures on later health problems [8]. The study invites participation for "well-child" visits at ages 2, 4, 6, 9, 12, 15, 18, 24 months and then every 12 months. We considered well-child visits after age 2 months because newborn infants may have numerous unscheduled visits and hence miss the well-child visit. Well-child visits tend not to occur before the child reaches a given age as children are unable to receive vaccinations and physicians cannot bill early visits. As a result, the left sides of bins were fixed at 5% of the gap between successive well-child visits, and the right side of bins were varied from 1% to 95% of the gap. Parents also brought their children to clinics for as-needed "sick visits". This dataset required careful consideration of the at-risk sets for each bin to account for censoring. A child was eligible for a bin if they entered the study before the left cut-point of the bin and were censored after the right cut-point of the bin. The at-risk sets were used to calculate the proportions of individuals with 0 and >1 visits per bin.

The years of recruitment were from 2008 to 2015. Individuals were excluded if they never had a well-child visit, had their first well-child visit after 24 months of age, or if their study entry and censoring dates could not be determined. The final dataset comprised 6,470 individuals with a median follow-up of 5.32 years. Visits from all 6,470 individuals were included in bin calculations, and Figure 3.13 illustrates the age at each visit for a randomly selected subset of 20 individuals.

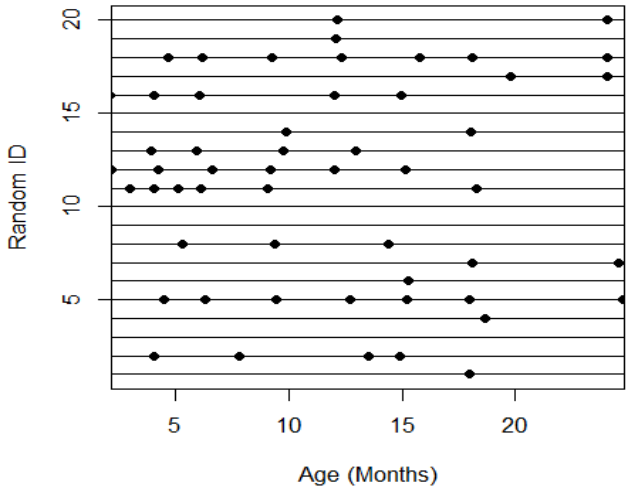

Figure 3.13: The age at visit (months) for a random subset of 20 individuals from the TARGet Kids! cohort.

We computed the AUC_{OBS}, AUC_{MLE} and AUC0. The likelihood function was maximized separately for each bin and assumed an independence structure for the missingness, sick visits, and well-child visit processes. Missing well-child visits were specified as Bernoulli random variables and the probability of missingness was allowed to differ across bins (π_j). Sick visits were assumed to follow a homogenous Poisson process within each bin, and rates were allowed to differ across bins (λ_j). Since most well-child visits occurred on or after the designated age, the timings of the j^{th} well-child visit ($\mathcal{T}_{ij}$ for subject i) were subtracted by the left cut-point of the j^{th} bin (i.e. $\mathcal{T}_{ij} - L_j$) and were modelled using the Log-normal(μ_j, σ_j) and Gamma(α_j, β_j) distributions.

As can be seen from Figure 3.14, the Log-normal distribution reflected the empirical densities more adequately than the Gamma distribution (other bins were similar), and modelling well-child visits using the Log-normal distribution yielded an AUC closer to the observed AUC (0.014 vs. 0.014) than the Gamma distribution (0.016 vs. 0.014) (Figure 3.15). Overall, the Log-normal distribution was a better fit and was used to estimate the likelihood-based AUC.

The mean proportions of individuals with 0 visits per bin were above 0.67 for all bin widths, and the mean proportions of individuals with >1 visit per bin were below 0.03 (Figure 3.15). The likelihood parameter estimates are summarized in Table 3.3. For the likelihood-based curve assuming no missingness, the estimated mean proportions of individuals with >1 visit per bin were below 0.07 and the AUC0 was 0.007 (Figure 3.15). This is consistent with repeated measures data since the AUC0 was near 0 and visits appeared to be regular (Figure 3.13). The simulation results from Section 3.4 indicated that an AUC of 0.007 could be associated with bias; however, in this case the primary contributors to the observed irregularity were the variability in scheduled visit timings and missingness, not the rate of unscheduled visits. Overall analyses using the AUC suggest that the TARGet Kids! cohort should be treated as repeated measures subject to missing data. Missing data techniques such as multiple imputation [39] or inverse-probability weighting [40] should be used to model the outcome trajectory.

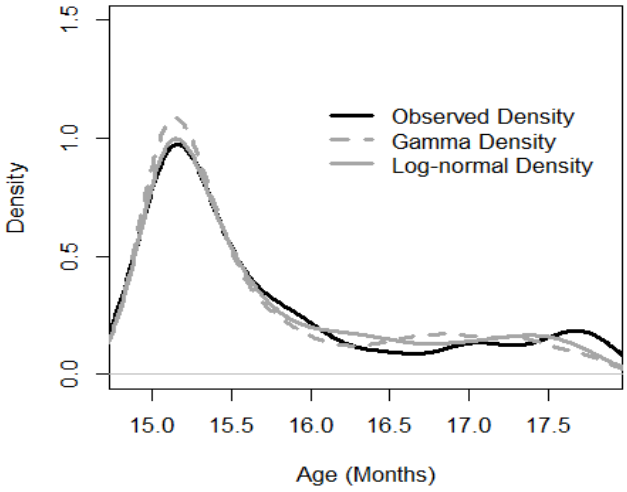

Figure 3.14: The likelihood-based densities using the Log-normal and Gamma distributions for well-child visits against the empirical density of visits for age 15 months for the TARGet Kids! cohort.

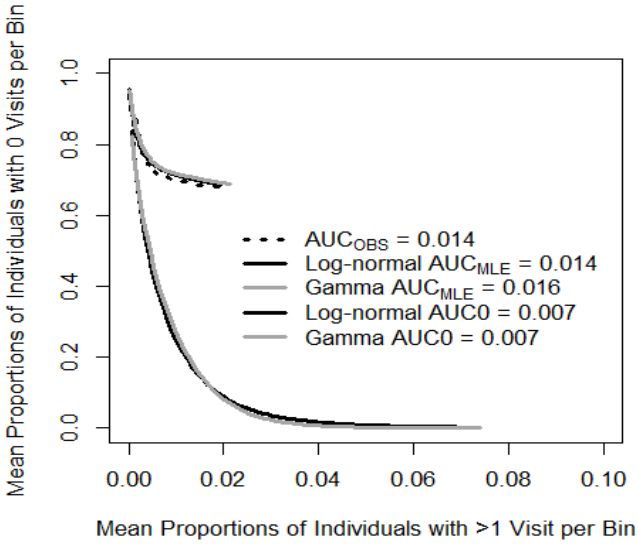

Figure 3.15: The observed AUC (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) for the TARGet Kids! cohort.

Well-child Age	**4 months**	**6 months**	**9 months**	**12 months**	**15 months**	**18 months**	**24 months**
μ_j	-0.999	-1.010	-0.537	-0.726	-0.673	-0.339	-0.025
σ_j	0.472	0.584	0.615	0.398	0.448	0.670	0.474
π_j	0.831	0.877	0.733	0.775	0.885	0.600	0.468
λ_j	0.047	0.029	0.023	0.030	0.020	0.010	0.016

Table 3.3: The likelihood-based parameter estimates for each well-child age corresponding to the Log-normal distribution for the timing of scheduled visits subtracted by the left cut-point of each bin (μ_j and σ_j), level of missingness (π_j), and the rate of unscheduled visits (λ_j) for the TARGet Kids! cohort.

3.6 Discussion

This paper proposes using the area under the curve (AUC) of the mean proportions of individuals with 0 vs. >1 visits per bin curve as a single score to quantify the extent of visit irregularity. We showed that the AUC increased with increasing irregularity and was invariant to sample size and the number of scheduled measurement occasions. Moreover, the likelihood-based AUC0 was invariant to missingness.

Irregular longitudinal data are frequently mishandled in practice and this can lead to biased results [34]. Two semi-parametric methods which incorporate the visit process into analyses on the longitudinal outcome are inverse-intensity weighted generalized estimating equations [20], and joint models of the outcome and visit processes [22]. Both methods are two-staged approaches wherein the visit process is initially modelled using regression models for recurrent events [16], and then generalized estimating equations are weighted by the inverse of each individual's estimated visit intensity, or the visit and outcome processes are jointly modelled using shared random effects. Assessing the extent of irregularity can help determine whether data should be treated as repeated measures (e.g. generalized estimating equations) or whether methods for irregular data should be considered (e.g. inverse-intensity weighted generalized estimating equations). Our previous paper addressed the lack of techniques for quantifying the extent of irregularity by creating bins across the study period and counting the proportion of individuals with 0, 1, and >1 visits per bin. This paper expands on our previous work and proposes the AUC as a measure for determining the extent of irregularity.

It is natural to wonder how large of an AUC is still tolerable for treating data as repeated measures. Just as with missing data, there is no threshold for deciding at what point irregularity is problematic and must be addressed to avoid biased results. With missing data, the severity of bias due to missingness depends on both the percentage of missingness and the missing data mechanism [33]; so with irregularity, bias depends on both the extent of irregularity and the relationship between the visit and outcome processes. The simulation results evaluating the relationship between the AUC and bias demonstrated that when the unscheduled visit process was not informative (i.e. no bias), the AUC assumed a range of values. When the unscheduled visit process was informative, the AUC increased with increasing bias in an exponential fashion. We highlighted potential thresholds for the AUC based on changes in bias: an AUC below 0.005 was considered to be small, and an AUC above 0.08 was deemed large in that context. However, certain visiting scenarios could yield different relationships between the AUC and bias, which in turn may lead to different AUC thresholds. The likelihood-based AUC can be helpful in determining how the data should be treated because it can provide important information concerning

the sources of irregularity. This was the case in the TARGet Kids! study where although an AUC of 0.007 could be tied to bias according to the simulation results, the rate of unscheduled visits was a minor contributor to the variability in visit timings, and thus the data was treated as repeated measures with missingness. As a result of the complex relationship between bias and the AUC (and other factors), we do not propose general thresholds for the AUC. The standard practice with missing data is to further investigate missingness (e.g. using logistic regression to identify predictors of missingness) within the context of the study and ultimately make an informed judgement call. We advocate the same approach to interpreting the AUC.

Estimating the likelihood-based AUC can be difficult when the standard deviation of scheduled visit timings is large. Large standard deviations make it difficult to distinguish between the scheduled and unscheduled visit processes. This was apparent in the case where unscheduled visits affected missingness; the correct likelihood-based AUC was overestimated for the larger value of the standard deviation because the parameter estimates of the unscheduled visit rate and shift parameter for the probability of missingness were inflated.

It would be ideal if the AUC was invariant to the follow-up length. In the case of no pre-specified visit times, the relationship between the AUC and gap time variance depended on the follow-up length; for larger variances, the AUC was larger for the longer follow-up time. This happened because visits are more irregular as the gap time variance increases, and thus a longer follow-up time would be needed to observe individuals with >1 visit. For smaller variances, the AUC was larger for the shorter follow-up time because the decrease in individuals with >1 visit was offset by the increase in individuals with 0 visits per bin. However, changes in AUC due to the follow-up length were small. Furthermore, most studies are designed to ensure that every individual has a measurement at baseline and then at least 1 follow-up visit, and thus force the probability of an individual having >1 visit over the course of the study to be 1.

The AUC could be explored in scenarios not covered in this paper. In the case of pre-specified visit times, most of the simulations were based on the assumption that the missingness, scheduled, and unscheduled visit processes were independent. We explored a specific violation of this assumption where unscheduled visits impacted missingness and we evaluated the consequences of specifying an independence structure for the likelihood-based AUC. In practice, there are other violations of the independence assumption. For example, future work could examine cases when missingness is impacted by delayed scheduled visits in the previous bin. Another scenario in which the AUC can be evaluated is when an unobserved process leads to an increased rate of unscheduled visits. Our simulations were based on the specification that the initial rate and increased rate of unscheduled visits were the same across individuals. Future work could examine models for such processes and explore the impact on the AUC of mistakenly assuming homogeneity in the unscheduled visit rate.

3.7 Conclusions

Irregular longitudinal data merits careful inspection to avoid biased conclusions on the longitudinal outcome. While exploring missing data is standard practice, apart from our previous paper there are currently no proposed measures for quantifying the extent of irregularity. This paper proposes using the AUC to fill this gap. The AUC is a single score which can be used to quantify the extent of irregularity and assess how closely visits resemble repeated measures. The AUC can improve the quality of statistical

analyses by guiding researchers in selecting the appropriate analytic outcome approach and minimize the potential for biased results.

Chapter 4

Guide to Exploring Irregularity

This chapter offers a guide for summarizing the extent of irregularity and discusses different considerations when characterizing the visit process, and selecting the modelling approach for the outcome. The following sections list the recommended steps in summarizing the extent of irregularity, and the final section reviews different types of research objectives and discusses how certain factors within each case can determine the suitability of a chosen modelling approach.

A simulated dataset is used to illustrate each step of the guide. The dataset has a sample size of 100 individuals with five scheduled measurement occasions at 2, 4, 6, 8, 10 months. The dataset consists of three independent processes corresponding to the timings of scheduled visits, missing scheduled visits, and unscheduled visits. For the purpose of this demonstration, it is assumed that there is no information to distinguish scheduled visits from unscheduled visits. The scheduled visit timings were normally distributed with means corresponding to the scheduled measurement occasions, and with a standard deviation of 0.3. Whether a scheduled visit was missing or not was specified as a Bernoulli distribution with probability of 0.2. Unscheduled visits were simulated from a homogeneous Poisson process with a rate of 0.3. The measures of irregularity will be used to assess whether this dataset should be treated as repeated measures.

4.1 Meet Team Who Conceptualized the Study

The first step to exploring visits is to meet with the team who conceptualized the study and gain familiarity with the clinical context and the study. This is important because understanding the research objectives and the study design can inform the specification of certain components of the measures of irregularity, and can ultimately provide a frame of reference for judging the measures. For example, if visits were intended to be repeated measures, then the measures corresponding to pre-specified visit times should be implemented, and the resulting values should be compared to values consistent with repeated measures (e.g. AUC of 0). Furthermore, these discussions can help gauge whether important predictors of visit intensity are captured in the dataset, and whether there could be latent predictors of visit intensity or missingness.

4.2 Plot Random Subset of Visits

After gaining familiarity with the study, the next step is to select a random subset of individuals from the dataset and plot their visit timings across the study period. This plot offers a snapshot of observed visit patterns and provides an insight into the extent of irregularity. This plot can also help determine the locations of bins in cases where there is no study protocol, or protocol is not adhered to.

4.2.1 Simulated Dataset

The visit timings for a random subset of 30 individuals from the simulated dataset are plotted across the duration of follow-up (Figure 4.1).

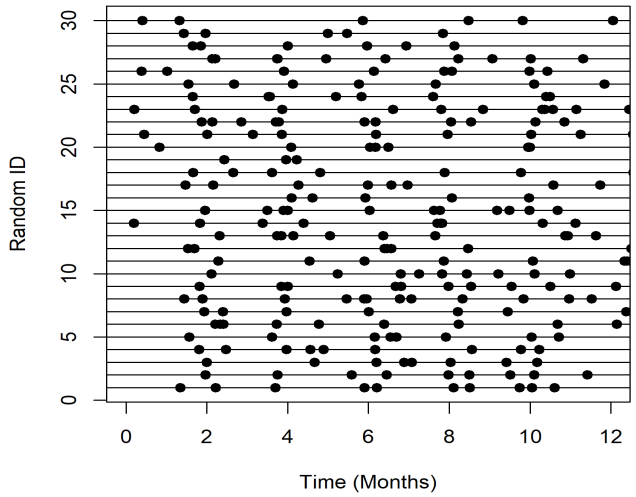

Figure 4.1: The visit timings for a random subset of 30 individuals from the simulated dataset.

This plot reveals that it is difficult to visually distinguish scheduled visits from unscheduled visits for some individuals. This can be attributed to multiple factors such as a high level of missingness (and most of the visits are unscheduled visits), or variability in scheduled visit timings (with some unscheduled visits). In either case, there is a potential for bias as missingness, the rate of unscheduled visits, and scheduled visit timings could be associated with the outcome process. In general, consistent deviations from protocol across individuals should be reported to the team who conceptualized the study as they can be indicative of an informative visit process.

4.3 Constructing Bins

Implementing the measures of irregularity begins by constructing bins across the study period. The study design and clinical context can help determine the type of bins (i.e. pre-specified visit times vs. no pre-specified visit times), and whether bins should have fixed or variable widths. In the TARGet Kids! study, it was noted that most children had their well-child visit once they reached a particular

age. This feature was incorporated into bin specifications by fixing the left cut-points at 5% of the gap between successive well-child ages, and varying the right cut-points. In the cSLE study, individuals were recommended to visit at least once every 6 months; however, it was suspected that visits occurred more frequently when disease activity worsened. As a result, the bin type corresponding to no pre-specified visit times was deemed suitable for this dataset, and the whole study period was split into adjacent and equally sized bins (with respect to time), and the number of bins was varied.

4.3.1 Simulated Dataset

To assess whether visits resemble repeated measures, bins will be centred around the scheduled measurement occasions and the left and right cut-points will be specified as a percentage of the gap between successive scheduled measurement occasions, and will be varied (i.e. the pre-specified visit times case).

4.4 Visual Measures of Irregularity

To visually assess the extent of irregularity, the mean proportions of individuals with 0, 1, and >1 visits per bin are plotted as a function of bin width. In Chapter 2, it was demonstrated that perfect repeated measures corresponds to every individual having 1 visit per bin, and missing scheduled visits lead to individuals with 0 visits per bin, whereas unscheduled visits lead to individuals with >1 visit per bin.

Evaluating the mean proportions can help determine whether observed visits have a repeated measures structure. For example, one way of doing this is to compare the mean proportions of individuals with 0 and >1 visits per bin at the bin width which yields the largest mean proportion of individuals with 1 visit per bin. Another way is to compare the mean proportions of individuals with 0 and 1 visits per bin at the widest bin width such that the mean proportion of individuals with >1 visit per bin is 0 (or at its lowest value). In the TARGet Kids! study, the mean proportions of individuals with 0 visits per bin were above 0.67 for all bin widths while the mean proportions of individuals with >1 visit per bin were below 0.03. These results suggested that visits were consistent with repeated measures subject to missingness, and were not indicative of irregularity.

4.4.1 Simulated Dataset

The mean proportions of individuals with 0, 1, and >1 visits per bin are plotted as a function of bin width (Figure 4.2).

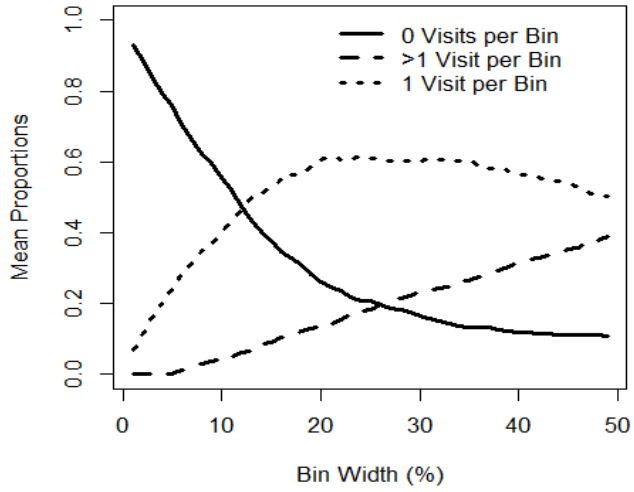

Figure 4.2: The mean proportions of individuals with 0, 1, and >1 visits per bin from the simulated dataset.

The largest mean proportion of individuals with 1 visit per bin is 0.61, and this occurs when the cut-points are at 24% of the gap between successive scheduled measurement occasions. At the same bin width, the mean proportions of individuals with 0 and >1 visits per bin are 0.21 and 0.18 respectively. Alternatively, the widest bins such that the mean proportion of individuals with >1 visit per bin is 0 occur when the cut-points are at 2% of the gap between successive scheduled measurement occasions. At this bin width, the mean proportions of individuals with 0 and 1 visits per bin are 0.88 and 0.12 respectively. For the widest bin widths (50% of the gap), the mean proportion of individuals with 0 visits per bin is 0.11 (the mean proportion of individuals with 1 and >1 visits per bin are 0.50 and 0.39 respectively). These results imply that there is a degree of irregularity in visit timings, and this could be explained by the presence of unscheduled visits, missing scheduled visits, and variability in scheduled visit timings.

4.5 Estimating the AUC

To quantify the extent of irregularity, the mean proportions of individuals with 0 visits per bin are plotted against the mean proportions of individuals with >1 visit per bin, and the AUC is estimated. In Chapter 3, it was shown that perfect repeated measures leads to an AUC of 0, and that the AUC increases as the level of missingness and the extent of irregularity increase. To negate the influence of missingness on the AUC, the likelihood function of visits is maximized and an AUC assuming no missingness is derived (AUC0). The team who conceptualized the study should be consulted when constructing the likelihood function because their expertise can be helpful when specifying the processes relating to scheduled visit timings, missingness, and unscheduled visits, and when characterizing the relationships between the processes (i.e. whether it is valid to assume the processes are mutually independent). For example, it is typically possible to establish whether scheduled visits timings will be symmetrically distributed around

the scheduled measurement occasions, or whether the distributions will be skewed. This was apparent in the TARGet Kids! study wherein the Gamma and Log Normal distributions were suitable candidates for the timings of well-child visits because certain features of the study (e.g. most well-child ages were vaccination ages) implied that the distributions would be right-skewed. In a similar fashion, missingness can be modelled as a function of covariates if it suspected that data are missing at random, and the rate of unscheduled visits can be modelled as a function of past outcomes to accommodate beliefs that the unscheduled visit process is informative.

4.5.1 Simulated Dataset

The likelihood function of visits was constructed by assuming that the scheduled visit timings were normally distributed with means corresponding to the scheduled measurement occasions, and that there was a common standard deviation σ across bins. The probability of missingness π was assumed to be the same across bins. Unscheduled visits were modelled using a homogenous Poisson process with rate λ. The three processes were assumed to be independent of each other.

Figure 4.3 displays the AUC estimated from the mean proportions of individuals with 0 vs. >1 visits per bin curve (AUC_{OBS}), and the likelihood-based AUCs (AUC_{MLE} and AUC0). Table 4.1 presents the parameter estimates.

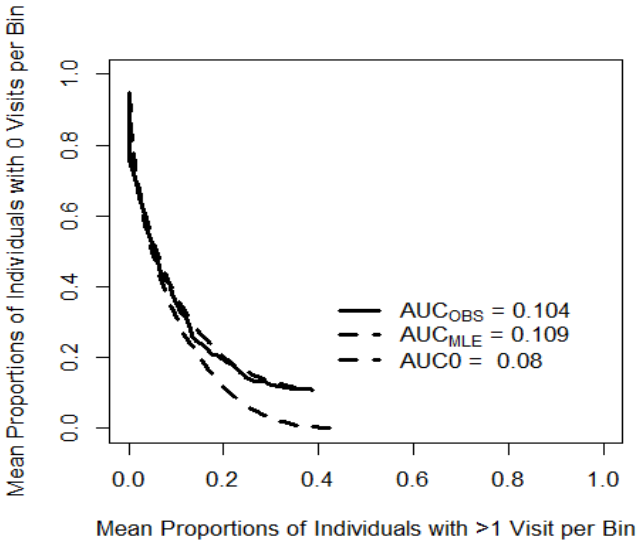

Figure 4.3: The AUC based on the mean proportions of individuals with 0 vs. >1 visits per bin (AUC_{OBS}), and the likelihood-based AUCs (AUC_{MLE} and AUC0) for the simulated dataset.

Parameter	Estimate
σ	0.304
π	0.194
λ	0.301

Table 4.1: The likelihood-based parameter estimates of the standard deviation of the scheduled visit timings (σ), the level of missingness (π), and the rate of unscheduled visits (λ).

The AUC0 is not near 0 (AUC0 = 0.08), and this supports the notion that there may be unscheduled visits and variability in scheduled visit timings as both contribute to an increased AUC. The likelihood-based parameter estimates as well as the visual measures of irregularity are consistent with this conclusion. In Section 3.4, an AUC above 0.08 was considered to be large in terms of the change in bias, and furthermore, the levels of bias were more pronounced when the rate of unscheduled visits was an important contributor to the extent of irregularity. These results reaffirm that a careful consideration of whether the data should be treated as repeated measures is required.

4.6 Determining the Importance of Irregularity

The decision of whether the observed irregularity is important enough to merit methods for irregular data can be informed by a thorough assessment of irregularity which includes summarizing the extent of irregularity, and evaluating predictors of visit intensity.

To perform a comprehensive summary of the extent of irregularity, the visual measures and the AUC should be interpreted together. This assessment can help with the decision of whether to use standard methods for longitudinal data (e.g. generalized estimating equations) or methods for irregular data (e.g. inverse-intensity weighted generalized estimating equations). In the TARGet Kids! study, all of the measures supported the conclusion that the dataset could be viewed as repeated measures with a substantial amount of missing data. These results were conveyed to the TARGet Kids! collaborators as modelling of a longitudinal outcome using standard methods for longitudinal data would likely be susceptible to bias. Future work may link to ICES data to determine whether visits were well-child visits or sick visits in order to improve understanding of the missingness rates.

Establishing whether there are predictors of visit intensity can help infer the relationship between the outcome and visit process. This is a necessary step in selecting a modelling approach for the outcome because if it was decided that the visit process was not VCAR, then standard methods for longitudinal data may not be valid. In addition, the underlying assumptions of irregular data methods such as inverse-intensity weighted generalized estimating equations may also be violated because it is impossible to empirically determine whether the visit process is VAR or VNAR. In the cSLE study, regression analyses indicated that disease activity was positively associated with visit intensity, which suggests that the visit process can at best be VAR. As a result, tailoring an appropriate modelling approach for the outcome requires careful assessment of the informativeness of the visit process.

Another important decision to be made is to judge how problematic the extent of irregularity is in terms of the potential for bias. This determination should be performed in conjunction with an evaluation of the informativeness of the visit process because a given degree of irregularity may be tolerable if the visit process was not informative, but may not be as acceptable if the visit process was informative.

4.6.1 Simulated Dataset

Although visits were intended to be repeated measures, all of the measures pointed towards an imbalance in visit patterns across individuals which was driven primarily by unscheduled visits and variability in scheduled visit timings. The team who conceptualized the study can play a crucial role in characterizing the visit process, and depending on how informative the visit process is judged to be, methods for irregular data may be appealing.

4.7 Modelling the Outcome

Summarizing the extent of irregularity and judging the visit process are crucial steps in selecting an appropriate modelling approach. There is no approach that can accommodate all visiting scenarios, and thus assumptions relating to the visit process need to be carefully evaluated. Table 4.2 summarizes the assumptions concerning the relationship between the outcome and visit process for standard methods for longitudinal data (mixed effects models and generalized estimating equations), and methods for irregular data (semi-parametric joint models and inverse-intensity weighted generalized estimating equations).

Modelling Approach	**Outcome and Visit Process Conditionally Independent Given:**				
			Previously Observed Covariates		
	Previously Observed Outcomes	**Previous Visits**	**Outcome Model**	**Visit Process Model**	**Random Effects**
Mixed Effects Models [10]	✓	✗	✓	✗	✗
Generalized Estimating Equations [9]	✗	✗	✓	✗	✗
Semi-Parametric Joint Models [22]	✗	✗	✓	✓	✓
Inverse-Intensity Weighted Generalized Estimating Equations [20]	✓	✓	✓	✓	✗

Table 4.2: The underlying assumptions concerning the relationship between the outcome and visit process for standard methods for longitudinal data (mixed effects models and generalized estimating equations), and methods for irregular data (semi-parametric joint models and inverse-intensity weighted generalized estimating equations) in terms of previously observed outcomes, visits, previously observed covariates in the outcome and visit process models, and random effects.

For example, Table 4.2 indicates that inverse-intensity weighted generalized estimating equations rely on the assumption that the outcome and visit process are conditionally independent given previously observed outcomes, previous visits, and previously observed covariates from the outcome and visit process models (i.e. the observed data by a given time t $F^{obs}(t)$).

The existence of latent variables which are predictive of visit intensity or missingness can also narrow the options for modelling the outcome. For example, if predictors of visit intensity are captured (and there are no latent predictors), then inverse-intensity weighted generalized estimating equations are valid, whereas if there are also latent predictors of visit intensity, then semi-parametric joint models may be valid provided that the latent predictors are not time-varying. In cases where visit intensity depends on a time-varying covariate (and possibly on time-invariant latent covariates), multiple outputation [51] can be a particularly useful option. Multiple outputation is a technique of sampling within clusters wherein the sizes of clusters are informative [52]. This concept is applied to irregular data where visits from each individual are randomly selected and discarded, and then the outcome is modelled using techniques for uncorrelated data. This process is repeated multiple times and the estimates from each dataset are averaged to yield a single value.

As previously mentioned, the research objectives can determine the choice of modelling approach. There are cases where the suitability of methods for standard longitudinal data (e.g. mixed effects models) as well as methods for irregular data (e.g. semi-parametric joint models) depend on the target of inference. The following sections discuss the different types of research objectives and comment on the validity of these methods.

4.7.1 Estimating the Overall Mean of the Outcome

There are studies which are interested in estimating the overall mean of an outcome over time. For example, Buzkova *et al* [35] reference a study where the primary research objective was to estimate the prevalence of pneumonia amongst Kenyan mothers with HIV-1. For these types of studies, if it suspected that there are covariates which are predictive of visit intensity, then methods for irregular data can be more effective than standard methods for longitudinal data in minimizing bias. Specifically, inverse-intensity weighted generalized estimating equations are recommended if there are no latent predictors of visit intensity.

4.7.2 Evaluating the Outcome Trend

Another common research objective is to evaluate the change in the outcome over time. In the cSLE study, one of the research objectives was to explore the disease trajectory of cSLE patients and how it evolved over time. In these cases, defining the modelling quantity as the change in the outcome from baseline ($Y_i(t) - Y_i(t_{i1})$) is a popular approach because it can simplify parameter estimation for methods which utilize random effects. To illustrate this, consider a random intercept model of the form $Y_i(t) = u_i + f(t)\beta + \epsilon_i(t)$ where $f(t)$ is some function of time t, and β is a scalar regression parameter. This model would transform into a model of the form:

$$Y_i(t) - Y_i(t_{i1}) = (f(t) - f(t_{i1}))\beta + (\epsilon_i(t) - \epsilon_i(t_{i1})) \tag{4.1}$$

which can restore the conditional independence between the outcome and visit process given the covariates if the random intercept is the only latent variable.

Inverse-intensity weighted generalized estimating equations can model the change in the outcome from baseline provided that the assumption of visiting by time t is conditionally independent of the

change from baseline in the outcome process at time t given the observed data by time t ($F^{obs}(t)$) holds.

4.7.3 Assessing Covariate Effects

Many studies are interested in evaluating the effects of covariates, whether it be on the overall mean of the outcome, or on the outcome trajectory. Neuhaus *et al* [53] compare the performances of mixed effects models to mis-specified irregular data methods (inverse-intensity weighted generalized estimating equations and semi-parametric joint models) with regards to the estimation of covariate effects across a range of visiting scenarios. The visiting scenarios were classified according to the ratio of scheduled visits to unscheduled visits, and the magnitude of the dependency of missingness and visit intensity on the outcome. Their simulation studies considered an outcome which was derived as a function of fixed and random effects, and the results indicated that mixed effects models showed negligible bias when covariates were not associated with the random effects, and low bias when covariates were associated with the random effects. In comparison, the mis-specified methods for irregular data showed greater bias. It was noted that the presence of scheduled visits was protective from bias only for the mixed effects models.

Choosing a modelling approach to assess covariate effects can hinge on whether the covariates are shared with the visit process. If the covariates are predictive of visit intensity, then irregular data methods may be preferred to standard methods for longitudinal data. If the covariates are not shared with the visit process, then the simulation results of Neuhaus *et al* [53] can be used as a guide to determine whether mixed effects models or methods for irregular data should be used to model the outcome.

Chapter 5

A Complete Demonstration: The STAR*D Study

This chapter applies the measures of irregularity from this thesis to the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study [54]. The STAR*D study is the largest randomized clinical trial on patients suffering from major depression [55]. This chapter focuses on the first phase of the study which had a common set of scheduled measurement occasions for all individuals, and thus interest lies in determining whether visits can be treated as repeated measures. This is followed by a demonstration on how to select the appropriate modelling approach for the study outcome, and how to interpret the resulting parameter estimates.

5.1 Study Background

The first phase of this study was over a 12 week period in which Citalopram (a selective serotonin reuptake inhibitor (SSRI) antidepressant) was administered. During this period, individuals were instructed to answer a questionnaire called the Quick Inventory of Depressive Symptomatology (QIDS) [56]. The QIDS questionnaire was designed to assess the severity of depressive symptoms by asking questions relating to multiple items such as sleep and appetite patterns. The QIDS score ranges from a total of 0 to 27 with higher scores indicating an increased severity of depressive symptoms. The target of inference of this chapter is to evaluate the mean QIDS score over the first 12 weeks of the trial.

The study protocol pre-specified a common set of scheduled measurement occasions at weeks 2, 4, 6, 9, 12 post-baseline during which individuals had their QIDS score recorded; however, there were individuals who missed scheduled visits, and had unscheduled visits. The QIDS score was available on 4,027 individuals.

5.2 Measures of Irregularity

To explore the extent of irregularity in this dataset, the visual measures of irregularity and a likelihood-based AUC were assessed.

5.2.1 Plot Random Subset of Visits

The visit timings for a random subset of 30 individuals from the first phase (12 weeks post-baseline) are presented in Figure 5.1.

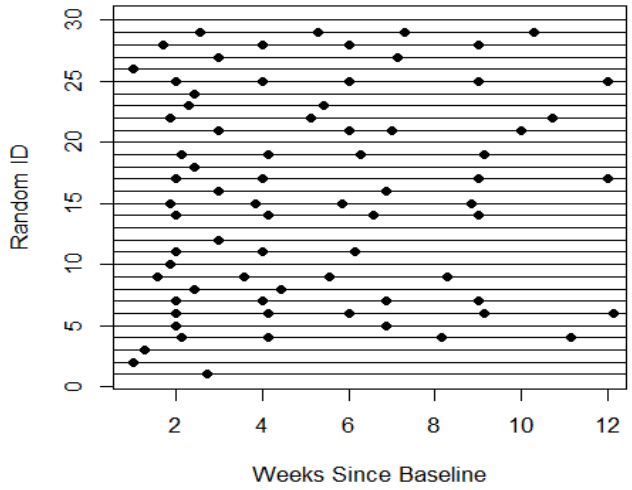

Figure 5.1: The visit timings (weeks since baseline) for a random subset of 30 individuals from the STAR*D study.

The majority of individuals had 1 visit at each scheduled measurement occasion; however, there was variability in visit timings across individuals. In addition, there were also individuals with 0 visits at each scheduled measurement occasion, and the proportion of individuals with 0 visits increased for the scheduled measurement occasions later on in follow-up.

5.2.2 Constructing Bins

Since there were five common scheduled measurement occasions occurring at 2, 4, 6, 9, 12 weeks post-baseline, bins were defined according to the pre-specified visit times case. To reflect the fact that most individuals had their scheduled visit on or after the exact scheduled measurement occasions, the left cut-points of each bin were fixed at 3 days before the scheduled measurement occasion, and the right cut-points were varied by the percentage of the gap between successive scheduled measurement occasions up to the largest possible values such that the bins were non-overlapping. For the bins which had the most number of days at their widest specification (i.e. 6, 9, 12 weeks), the largest right cut-points were at 76% of the gap between successive scheduled measurement occasions.

5.2.3 Visual Measures of Irregularity

Evaluating the mean proportions of individuals with 0, 1, and >1 visits per bin can help determine whether observed visits are consistent with repeated measures. Figure 5.2 displays the mean proportions of individuals with 0, 1, and >1 visits per bin as a function of bin width.

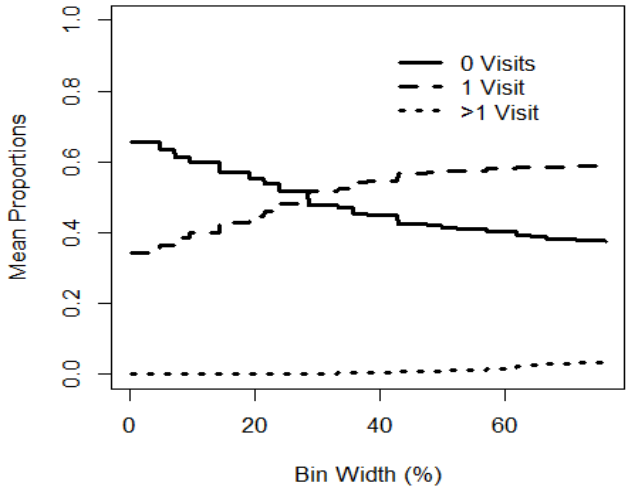

Figure 5.2: The mean proportions of individuals with 0, 1, and >1 visits per bin from the STAR*D study.

The mean proportions of individuals with 0 visits per bin were above 0.37 for all bin widths while the mean proportions of individuals with >1 visit per bin were below 0.04. The bin width which yielded the largest mean proportion of individuals with 1 visit per bin occurred when the right cut-points were at 76% of the gap between successive scheduled measurement occasions. At this bin width, the mean proportion of individuals with 1 visit per bin was 0.59, and the mean proportions of individuals with 0 and >1 visits per bin were 0.37 and 0.04 respectively.

These values supported the suggestion that most individuals had 1 visit per bin (with some variability in scheduled visit timings), and that some individuals missed their scheduled visits. Overall, the visual measures implied that the data could be viewed as repeated measures with missingness.

5.2.4 Estimating the AUC

To quantify the extent of irregularity, the AUC_{OBS}, AUC_{MLE} and AUC0 were estimated. The likelihood function of visits was specified in the widest possible bins. Visit timings were modelled in terms of days, and were treated as discrete random variables under the assumption that at most 1 visit could occur in a day.

To formulate the likelihood function for the i^{th} individual ($i = 1$ to $4,027$) in the j^{th} bin ($j = 1$ to 5), let M_j denote the total number of days in the j^{th} bin. Let the observed visit times for the i^{th} individual in the j^{th} bin be $t_{ij1}, \ldots, t_{ijn_{ij}}$. It was assumed that unscheduled visits could only occur after the day on which a scheduled visit occurred. Let the probability of a scheduled visit occurring at day k' ($k' = 1, \ldots, M_j$) be $w_{jk'}$. This probability was estimated as the empirical probability that an individual's first visit within bin j was observed at day k'. The probability of missing a scheduled visit (π_j) was estimated as the empirical probability of an individual having 0 visits in the bin. The total number of unscheduled visits for the i^{th} individual in the j^{th} bin was modelled using a Binomial distribution with probability λ_j and the number of trials being the number of days after the day on

which the scheduled visit occurred in the respective bin (i.e. $R_j - t_{ij1}$).

The likelihood function was maximized separately for each bin with the parameters corresponding to scheduled visit timings, missing scheduled visits, and unscheduled visits being allowed to differ across bins.

The mean proportions of individuals with 0 visits per bin were plotted against the mean proportions of individuals with >1 visit per bin, and the AUC_{OBS}, AUC_{MLE} and AUC0 are displayed in Figure 5.3.

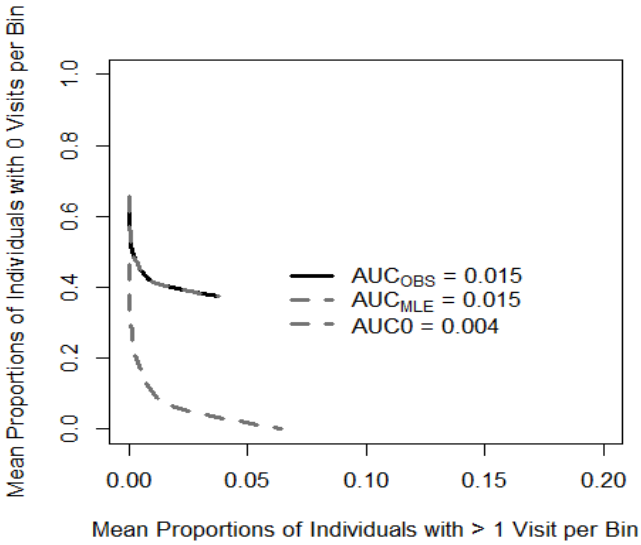

Figure 5.3: The observed AUC (AUC_{OBS}) and the likelihood-based AUCs (AUC_{MLE} and AUC0) for the STAR*D study.

The mean proportions of individuals with 0 visits per bin were above 0.37 for the observed and likelihood-based curves. For the likelihood-based curve assuming no missingness, the estimated mean proportions of individuals with >1 visit per bin were below 0.07, and the AUC0 was 0.004 (Figure 5.3). In Section 3.4, an AUC of 0.004 fell in the range of slower initial increase in mean bias. The likelihood-based parameter estimates corresponding to missingness, unscheduled visits, and scheduled visit timings are presented in Tables 5.1 and 5.2.

Parameter	**Scheduled Measurement Occasion**				
	2 Weeks	**4 Weeks**	**6 Weeks**	**9 Weeks**	**12 Weeks**
π_j	0.213	0.312	0.331	0.445	0.570
λ_j	0.004	0.004	0.007	0.004	0.008

Table 5.1: The likelihood-based parameter estimates for each bin corresponding to the level of missingness (π_j), and the unscheduled visit process (λ_j) for the STAR*D study.

	Day in Bin (k')																				
Bin	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
1	0.04	0.05	0.09	0.40	0.09	0.06	0.05	0.04	0.04	0.05	0.05	0.02	0.01	0.01	-	-	-	-	-	-	-
2	0.02	0.05	0.08	0.28	0.08	0.06	0.04	0.04	0.04	0.06	0.07	0.02	0.01	0.01	-	-	-	-	-	-	-
3	0.02	0.04	0.06	0.23	0.06	0.06	0.04	0.04	0.05	0.06	0.07	0.02	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01
4	0.03	0.03	0.06	0.17	0.06	0.05	0.03	0.03	0.04	0.05	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01
5	0.02	0.03	0.04	0.12	0.04	0.03	0.02	0.02	0.03	0.04	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01

Table 5.2: The likelihood-based estimates of the probability of a scheduled visit occurring on each day for all bins given that at least 1 visit was observed in the bin for the STAR*D study.

The estimated probability of missingness (π_j) increased across the duration of follow-up (Table 5.1). The estimated probabilities of a scheduled visit occurring at the exact scheduled measurement occasions (i.e. $k' = 4$) were smaller for the scheduled measurement occasions later on in follow-up, which suggested that the degree of variability in scheduled visit timings increased across the duration of follow-up (Table 5.2).

The AUC0 and the likelihood-based parameter estimates of the visit process components supported the conclusion that the observed visits resembled repeated measures data subject to missing scheduled visits and variability in scheduled visit timings. This was consistent with the assessment of the visual measures of irregularity.

5.2.5 Modelling the Visit Process

Selecting a valid modelling approach for the outcome requires a careful assessment of the underlying assumptions concerning the relationship between the outcome and visit process. It is important to explore potential predictors of visit intensity as standard methods for longitudinal data could be invalid if such predictors were identified. Predictors of visit intensity were identified by fitting a semi-parametric Cox proportional hazards regression model using the Andersen-Gill formulation. Available baseline individual-level characteristics included: interview age (years), sex, marital status, residential information (rental or owned, number of relatives or friends living at residence (reference category is living alone)), economic status (employment status (reference category is employed), monthly household income, and unemployment benefits received), and years of formal education.

The QIDS scores in the visit model were lagged by 1 visit for each individual. Model selection was based on fitting a regression model with all available predictors, and subsequently retaining predictors with P-Values < 0.05 in the final model. Analysis used the "coxph" function in R version 3.1.0 with cluster-robust standard errors [48]. The QIDS score at the previous visit was raised to the third power to satisfy the proportional hazards assumption of the Andersen-Gill model. Table 5.3 presents the results from the model which included all available predictors and the cubed QIDS score at the previous visit.

Characteristic	Time-Varying	Hazard Ratio	95% Hazard Ratio Confidence Limits	P-Value
Cubed QIDS Score at Previous Visit (75^{th} vs. 25^{th} Percentiles)	Yes	1.164	(1.141, 1.187)	<0.0001
Monthly Household Income ($10000)	No	1.010	(0.975, 1.046)	0.571
Unemployment Benefits (Yes/No)	No	1.032	(0.977, 1.090)	0.257
Employment Status	No			
Missing	-	0.793	(0.665, 0.946)	0.010
Unemployed	-	0.980	(0.955, 1.004)	0.103
Retired	-	1.026	(0.978, 1.075)	0.298
Student Status (Yes/No)	No	1.019	(0.986, 1.053)	0.265
Number of People in Household	No			
Missing	-	1.237	(0.993, 1.541)	0.057
2 People in Household	-	0.998	(0.968, 1.030)	0.920
>2 People in Household	-	0.968	(0.937, 1.000)	0.048
Marital Status (Yes/No)	No	1.023	(0.996, 1.050)	0.093
Rental Status (Yes/No)	No	0.986	(0.963, 1.010)	0.261
Sex (Male/Female)	No	0.957	(0.936, 0.978)	<0.0001
Age at Baseline (Decades)	No	0.996	(0.986, 1.006)	0.446
Formal Education (Decades)	No	1.062	(1.026, 1.099)	0.001

Table 5.3: The visit process modelling results including all predictors and the cubed QIDS score at the previous visit for the STAR*D study.

The final model included the cubed QIDS score at the previous visit, employment status, years of formal education, and sex as significant predictors of visit intensity. Table 5.4 presents the final model summary.

Characteristic	Time-Varying	Hazard Ratio	95% Hazard Ratio Confidence Limits	P-Value
Cubed QIDS Score at Previous Visit (75^{th} vs. 25^{th} Percentiles)	Yes	1.163	(1.141, 1.185)	<0.0001
Employment Status	No			
Missing	-	0.750	(0.616, 0.911)	0.004
Unemployed	-	0.979	(0.957, 1.002)	0.073
Retired	-	1.031	(0.989, 1.075)	0.146
Formal Education (Decades)	No	1.083	(1.049, 1.118)	<0.0001
Sex (Male/Female)	No	0.958	(0.938, 0.978)	<0.0001

Table 5.4: The visit process model for the STAR*D study.

An increase in the cubed QIDS score at the previous visit was associated with more frequent visits. The hazard ratio comparing the 75^{th} and 25^{th} percentiles for the QIDs score (Q3 = 15 and Q1 = 6) at the previous visit was 1.163 (95% confidence interval: 1.141-1.185). Individuals who had a missing

employment status visited less frequently compared to those who were employed (hazard ratio = 0.750, 95% confidence interval: 0.616-0.911), and individuals with more years of formal education visited more frequently (hazard ratio = 1.083, 95% confidence interval: 1.049-1.118). Males visited less frequently compared to females (hazard ratio = 0.958, 95% confidence interval: 0.938-0.978).

The following sections discuss how to model the study outcome based on the assessments of the measures of irregularity and the visit process modelling results.

5.3 Modelling the Outcome

The guide to exploring irregularity presented in Chapter 4 of this thesis discussed scenarios where standard methods for longitudinal data and specialized methods for irregular data are valid. The suitability of a chosen approach hinges on the relationship between the outcome and visit processes. Table 4.2 classifies this relationship according to previously observed outcomes, visits, covariates, and random effects. The exploratory analyses on visits can help with the judgement of this relationship. The visual measures of irregularity and the AUC were consistent with the notion that the data can be viewed as repeated measures with missingness and variable scheduled visit timings. The modelling of the visit process identified an association between visit intensity and the cubed QIDS score at the previous visit as well as employment status, years of formal education, and sex, suggesting that visits can at best be VAR. This rendered likelihood-based methods (e.g. standard mixed effects model) invalid as visit intensity was associated with covariates which were not intended to be included in the outcome model. The likelihood-based AUC analyses indicated that the main sources of irregularity were missingness and variable visit timings, and not the rate of unscheduled visits. As a result, missing data techniques such as multiple imputation and inverse-probability weighted generalized estimating equations were considered viable options under the assumption that the data were missing at random or missing completely at random. Multiple imputation creates multiple datasets in which the missing values are imputed by sampling from an estimated conditional joint distribution of the missing data, and then the target parameters are estimated by aggregating the dataset-specific estimates using Rubin's rule [39]. With inverse-probability weighted generalized estimating equations, individuals are weighted by the inverse of the estimated probability of being observed at each occasion as a function of the available predictors [40].

The data could also be viewed through the lens of irregular longitudinal data as repeated measures with missingness can be considered as a special case of irregularity. Specialized methods for irregular longitudinal data such as inverse-intensity weighted generalized estimating equations can account for predictors of visit intensity, and can analyze the outcome in the presence of missingness. When considering modelling approaches for evaluating the trend in the mean outcome over time as described in Section 4.7.2, the use of inverse-intensity weighted generalized estimating equations is valid provided that there are no latent predictors of visit intensity that are correlated with the outcome [15]. Semi-parametric joint models could not be used to evaluate the change in the mean QIDS score over time as the parameter representing the trend over time would be treated as a nuisance parameter and would not be estimated.

In this section, inverse-intensity weighted generalized estimating equations was deemed a suitable approach to model the mean QIDS score over time with the assumption of no latent visit predictors. The resulting parameter estimates were compared to estimates derived from: 1) multiple imputation, 2) inverse-probability weighted generalized estimating equations, and 3) unadjusted generalized estimating

equations with an unstructured working correlation structure. The multiple imputation approach was implemented by generating five imputed datasets where sex, years of formal education, and employment status were included as linear predictors of missingness. Unadjusted generalized estimating equations were applied to each imputed dataset and Rubin's rule was used to pool the estimates. The inverse-probability weights were estimated by modelling the probability of being observed at each occasion using logistic regression which was adjusted for the cubed QIDS score at the previous visit, sex, years of formal education, and employment status. The logistic regression model was fit on a reduced dataset in which individuals were artificially censored at their first missing scheduled visit. Unscheduled visits occurring before the first missing scheduled visit were utilized in the logistic regression model to help estimate the probabilities of being observed at each occasion. Figure 5.4 displays the QIDS scores over time across all individuals with an estimated Loess curve.

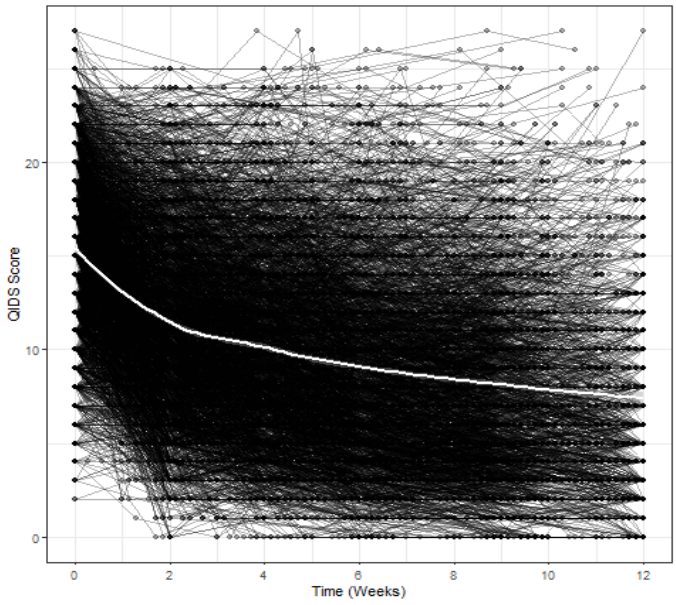

Figure 5.4: The QIDS scores over time across all individuals with an estimated Loess curve.

The estimated Loess curve showed a negative trend in the QIDS score over time which was not linear. As a result, the outcome model for all approaches implemented a square-root transformation to time and was specified as $E(Y_i(t)) = \beta_0 + \beta_1 t^{1/2}$.

These analyses were performed using R version 3.1.0. The inverse-intensity weighted generalized estimating equations approach was implemented using the "iiwgee" function [43]. Multiple imputation was performed using the "mice" function [57]. The inverse-probability weighted generalized estimating equations approach and the unadjusted generalized estimating equations regression models were implemented using the "geeglm" function [58]. The results of modelling the mean QIDS score using these approaches are presented in Table 5.5:

Modelling Approach	β_0 **Estimate (95% C.I)**	β_1 **Estimate (95% C.I)**
Inverse-Intensity Weighting	15.03 (14.89, 15.17)	-2.46 (-2.52, -2.41)
Inverse-Probability Weighting	15.04 (14.88, 15.19)	-2.41 (-2.47, -2.34)
Multiple Imputation	15.23 (15.09, 15.36)	-2.29 (-2.34, -2.24)
Unadjusted	15.13 (14.99, 15.26)	-2.22 (-2.28, -2.16)

Table 5.5: The modelling results for the STAR*D study.

All of the approaches suggested that the mean QIDS score decreased over time (β_1: P-Values < 0.0001). The unadjusted generalized estimating equations approach yielded a significantly larger estimate of β_1 compared to the results derived from using inverse-intensity weighting and inverse-probability weighting (Table 5.5), and the overall lowest estimate in absolute terms. Multiple imputation can lead to bias when there is notable missingness as it can be difficult for the imputation model to describe the joint distribution of the missing variables [59]. The rate of decline in the mean QIDS score over time was underestimated by the unadjusted generalized estimating equations approach because individuals with larger QIDS scores visited more frequently (Table 5.4), and the failure to account for this resulted in an over-representation of visits with larger QIDS scores in the analyses. The parameter estimates obtained from using inverse-intensity weighting and inverse-probability weighting were similar as the data resembled repeated measures with missingness, and the rate of unscheduled visits was not an important contributor to the extent of irregularity. The missingness model suggested that individuals with larger QIDS scores were less likely to miss scheduled visits, and this was consistent with the visit process model conclusion that these individuals visited more frequently. Although both methods were considered valid approaches to model the mean QIDS score over time, inverse-intensity weighting was preferred because the entire data was utilized, whereas the inverse-probability weighting approach was based on artificially censored data which risked a loss in efficiency [60] as information for individuals with intermittently missing data was discarded.

5.4 Conclusions

This chapter has demonstrated the importance of using the measures of irregularity in this thesis to thoroughly explore visits, and to subsequently determine a valid approach for modelling a longitudinal outcome which can minimize the potential for biased results. The visual measures of irregularity and the likelihood-based AUC indicated that there was variability in scheduled visit timings as well as missing scheduled visits which were increasingly prevalent later on in follow-up. Modelling the visit process indicated that the QIDS score at the previous visit, employment status, years of formal education, and sex were associated with visit intensity. This was important because it ruled out certain modelling approaches such as standard mixed effects models. Neglecting the visit process and modelling the outcome using unadjusted generalized estimating equations underestimated the rate of decline in the mean QIDS score over time, whereas using inverse-intensity weighting corrected this issue.

Chapter 6

Discussion

The measures of irregularity performed well relative to the properties we desired. Specifically, the AUC increased with increasing irregularity and was invariant to sample size and the number of scheduled measurement occasions, and the AUC0 was invariant to missingness.

We have proposed two measures for quantifying the extent of irregularity. In Chapter 2, we compared the proportions of individuals with 0, 1, and >1 visits per bin over the study period. This is an intuitive approach to assessing visit patterns across individuals, and it is straightforward to implement. However, it can be hard to summarize the mean proportions of individuals with 0, 1, and >1 visits per bin succinctly, and the choice of bin width can be subjective. In Chapter 3, we proposed the AUC estimator as a single measure of the extent of irregularity, evaluated its properties in different visiting scenarios, and provided benchmarks to help interpret the AUC. To perform a comprehensive assessment of the extent of irregularity, we recommend evaluating the mean proportions of individuals with 0, 1, and >1 visits per bin in conjunction with estimating the AUC as both measures reveal different aspects of visits. For example, to judge the proximity of data to repeated measures, the mean proportions of individuals with 0, 1, and >1 visits per bin obtained at the bin width which maximizes the mean proportions of individuals with 1 visit per bin can be compared to values which would be apparent if data were perfect repeated measures (i.e. whether the mean proportions of individuals with 0 and >1 visits per bin are near 0). The AUC0 can supplement this analysis by quantifying the extent of irregularity not influenced by the level of missingness.

It can be difficult to determine the magnitude of the AUC for concluding that data should not be treated as repeated measures. In other areas of statistics, there are recommended thresholds for interpreting a measure which have become general consensus (e.g. a VIF above 4 is taken to indicate a problematic level of multi-collinearity). In the simulation study exploring the relationship between bias and the AUC, we identified potential thresholds for the AUC based on the rate of change in bias where an AUC below 0.005 was deemed small, and an AUC above 0.08 was considered large. These thresholds were specific to this simulation setup and the parameter values considered, and should therefore be referenced with caution. We do not propose general thresholds for interpreting the AUC because thresholds can be arbitrary, and the implications of the AUC depend on the informativeness of the visit process. Similarly, there are no thresholds for interpreting the level of missingness as its implication depends on the missing data mechanism. For example, a level of missingness of 20% has different implications if the data are missing completely at random or missing not at random. If the data are missing completely at

random, then the available data can be thought of as a random sample of the target population. On the other hand, if the data are missing not at random, then the level of missingness is problematic as the available data are not representative of the target population, and modelling the outcome will yield biased estimates unless the missing data are properly addressed. This logic applies to irregular data, and thus the extent of irregularity should be judged in tandem with the informativeness of the visit process.

It may be appealing to test whether visits resemble repeated measures (i.e. test if the AUC is significantly different from 0) to circumvent the subjectivity of judging the AUC on its own. Furthermore, point estimates are reported along with some measure of variability when performing inference. However, this would not be appropriate for the AUC as it is intended to be a purely descriptive measure of the extent of irregularity for the data at hand analogous to the proportion of missing values; the AUC is either 0 or it is not. This rationale is why confidence intervals are not calculated for other descriptive measures such as the proportion of missingness or the I-squared statistic. Just as with the proportion of missing values, an informed judgement call of how large an AUC is still tolerable for treating the data as repeated measures (i.e. how close the AUC is to 0) is ultimately unavoidable. An important caveat is the AUC0 which has an element of potential uncertainty introduced through likelihood-based estimation of the visit process parameters. In this case, it could be of interest to estimate confidence intervals for the visit process parameters using asymptotic theory, or utilize methods such as bootstrapping to derive confidence intervals when closed-form expressions are not readily available.

The AUC is a succinct measure for quantifying the extent of irregularity, but there is a range of AUC values which can be ambiguous. For example, the meaning of an AUC of 0.1 is not immediately obvious in terms of whether data should be treated as repeated measures or not. We established that an AUC of 0 indicates perfect repeated measures, and that the AUC approaches 0.5 as the extent of irregularity increases. When there are pre-specified visit times, our simulations evaluated the AUC across different values of the standard deviation of scheduled visit timings, probability of missingness, and rate of unscheduled visits, and the results highlighted a plausible range of AUC values. Furthermore, we showed that if visits come from a Poisson process, the resulting AUC is 0.25. In general, our simulations were based on a range of visiting scenarios and different parameter specifications, and the results provide a point of reference for interpreting the AUC. The interpretability of the AUC could be improved by re-scaling it to an intuitive range. For example, the I-squared statistic and the intraclass correlation (ICC) are instinctively judged because they range from 0 to 1 and represent proportions. Although the AUC ranges from 0 to 0.5, most repeated measures scenarios tend to yield low AUC values. This is an area for future research. A potential approach to rescale the AUC to an intuitive range is to apply a transformation. For example, a log-based transformation such as $100(1 - \frac{\log(4(0.5-\text{AUC}))}{\log(2)})$ will become 0 if the data is perfect repeated measures (i.e. an AUC of 0), and becomes 100 if the visits are realizations from a Poisson process (i.e. an AUC of 0.25).

An important consideration when using the AUC to inform the modelling approach is the ratio of scheduled visits to unscheduled visits. There are cases when standard methods for longitudinal data are preferable to specialized methods for irregular data. Neuhaus *et al* [53] noted that fitting standard mixed effects models using maximum likelihood estimation when there are more scheduled visits than unscheduled visits can yield negligible bias in covariates not associated with the random effects and some bias in covariates associated with the random effects, whereas the specialized methods for irregular data will yield estimates with greater bias when mis-specified. However, their simulation results demonstrated that the level of bias when using mixed effects models (as well as specialized methods for irregular data)

does increase as the extent of irregularity increases. Therefore, we advocate for a thorough documentation of the extent of irregularity and a careful consideration of features of the visit process.

Study designs can have an impact on the extent of visit irregularity and the AUC. We recommend a proactive approach to minimizing the potential for irregular visit times as much as possible. For example, strategies can be implemented to make it more convenient for individuals to adhere to protocol (i.e. ensure regular visits) such as using electronic means (e.g. emails, text messages etc.) to remind them of their next visit time. In the TARGet Kids! study, using multiple forms of reminders might have improved the levels of missingness. Another approach is to obtain electronic information relating to the outcome or other individual characteristics relevant to the study if possible (e.g. electronic medical records). To address the potential for problematic levels of deviations from protocol, a monitoring system can be put in place to highlight individuals who are overdue for a visit. These individuals can then be contacted. For example, if it was discovered that an individual switched providers, the new providers could be approached and asked for both their and the patient's consent to share their chart information. In the cSLE study, visits were recommended to occur at least once every 6 months; however, there were individuals who did not visit for long periods of time (e.g. 3 years), and then resumed their visits. The problem with these visits is that they are likely to be informative, and thus the question arises of whether it is better to artificially censor these individuals (and discard information), or do nothing. Future cSLE studies could address this issue by surveying the individuals who had large gaps in their visits and identify the reasons for these gaps, and whether anything can be done to ensure that large gaps can be minimized in the future (e.g. phone calls to confirm drop-out, ask if they moved to a different provider etc.).

With the TARGet Kids! cohort, the lack of accurate information characterizing whether a visit was a well-child visit or a sick visit made it more difficult to estimate the AUC. If visits were labeled accurately, the likelihood function could incorporate this information and estimate a more robust AUC with respect to the standard deviation of scheduled visit timings. Specifically, the adverse impact on the AUC of large standard deviations in scheduled visit timings (which can make it difficult to distinguish between scheduled visits and unscheduled visits) would be alleviated. The TARGet Kids! data will be linked to data from the Institute for Clinical Evaluative Sciences (ICES), and so billing information can be used to record whether a visit was a well-child visit or a sick visit. In general, ensuring the accuracy of visit labels would have a positive impact on the estimation of the AUC, and thus improve the quality of exploratory analyses on the visit process.

This thesis has addressed the lack of measures for quantifying the extent of irregularity by proposing visual measures and the AUC, but there is more work that could be done. For example, the properties of the AUC were explored and its behaviour was examined under various visiting scenarios, but future work could explore likelihood-based estimation of the AUC where the unscheduled visit process is modelled semi-parametrically to reflect the fact that recurrent event models are typically specified semi-parametrically. Furthermore, improving the uptake of new statistical methods requires the availability of code [61]. R code for implementing the visual measures of irregularity from Chapter 2, and estimating the AUC are provided in Appendix C along with a worked example. The R functions from Appendix C are available for statistical analyses through the R package "IrregLong" in CRAN [43].

6.1 Conclusions

The measures of irregularity proposed in this thesis can help improve statistical analyses. "Big Data" (i.e. high volume data which are aggregated from multiple sources) are increasingly available and can be helpful in answering research questions. However, Big Data present a set of challenges including potentially informative visit timings. The problem of irregular data is analogous to the well-recognized problem of missing data, but is often mishandled in practice. We aim to raise awareness on the importance of quantifying the extent of irregularity and assessing predictors of visit intensity to make an informed decision on the modelling approach. There are specialized methods for irregular data which can accommodate a broad range of visiting scenarios and can help minimize the potential for biased results. Overall, the measures proposed in this thesis can lead to an improved quality of statistical analyses within longitudinal studies.

Appendix A: R Code Chapter 2

The following R code can be used to plot visit timings for a random subset of n individuals and the mean proportions of individuals with 0, 1, and >1 visits per bin. There is also R code which can be used to model the visit process using the Andersen-Gill formulation [18], to estimate weights which are the inverse of an individual's estimated visit intensity, and to test the proportional hazards assumption of the model. To implement this model using the "coxph" function [48], the data must be formatted into the counting process structure (an individual's data is split into rows corresponding to risk intervals of event times) [62]. The variables age and age_stop are used as the time variables in this code.

```
##Using the IrregLong Package to Generate Visual Measures

library(IrregLong)
library(Survival)

##Plot Visit Timings for a Random Subset of n Individuals

abacus.plot(n,time,id,data,tmin,tmax,xlab.abacus="Time",ylab.abacus="Subject", pch.abacus=16,col.abacus=1)

##Plot Mean Proportions of Individuals with 0, 1, and >1 Visits per Bin

extent.of.irregularity(data,time="time",id="id",scheduledtimes=NULL, cutpoints=NULL,ncutpts=NULL
,maxfu=NULL, plot=FALSE,legendx=NULL,legendy=NULL, formula=NULL,tau=NULL)

##Modelling the Visit Process

##Create an "event" Indicator Representing When a Visit Occurred
data$event<-c(1)

##Visit Process Model
model1<-coxph(Surv(time_stop, time, event)~agedx+factor(eth)+..., data=data)
summary(model1)

##Weights
data$p1<-predict(model1,newdata=data, type="lp")
data$weights<-1/data$p1
summary(data$weights)
```

```
##Test PH Assumption
model11<-cox.zph(model1)
model11
```

Appendix B: Tables Chapter 3

The following tables summarize the results from the simulation studies performed in Chapter 3.

τ	n	Gap Time Variance	Mean AUC_{OBS}	Standard Error
15	30	1.683	0.073	0.014
		3.542	0.104	0.017
		5.598	0.132	0.018
		7.869	0.154	0.020
		10.380	0.174	0.020
	100	1.683	0.072	0.008
		3.542	0.102	0.009
		5.598	0.132	0.010
		7.869	0.156	0.011
		10.380	0.175	0.010
30	30	1.683	0.062	0.008
		3.542	0.103	0.011
		5.598	0.139	0.013
		7.869	0.169	0.015
		10.380	0.193	0.016
	100	1.683	0.061	0.004
		3.542	0.102	0.006
		5.598	0.139	0.008
		7.869	0.168	0.008
		10.380	0.193	0.008

Table B.1: The mean observed AUCs (AUC_{OBS}) for Log-normal gap times across sample size (n) and study duration (τ).

τ	n	Gap Time Variance	Mean AUC_{OBS}	Standard Error
15	30	4.000	0.131	0.019
		5.000	0.144	0.019
		6.667	0.164	0.020
		10.000	0.199	0.020
		20.000	0.259	0.020
	100	4.000	0.132	0.010
		5.000	0.144	0.011
		6.667	0.165	0.011
		10.000	0.198	0.011
		20.000	0.259	0.011
30	30	4.000	0.122	0.011
		5.000	0.137	0.011
		6.667	0.160	0.013
		10.000	0.200	0.014
		20.000	0.273	0.016
	100	4.000	0.122	0.006
		5.000	0.137	0.007
		6.667	0.160	0.007
		10.000	0.200	0.008
		20.000	0.272	0.009

Table B.2: The mean observed AUCs (AUC_{OBS}) for Gamma gap times across sample size (n) and study duration (τ).

σ	π	λ	**Mean AUC_{OBS}**	**Standard Error**	**Mean AUC_{MLE}**	**Standard Error**	**Mean AUC0**	**Standard Error**
0.1	0.1	0.10	0.023	0.008	0.023	0.006	0.011	0.003
		0.15	0.033	0.011	0.032	0.008	0.015	0.003
		0.20	0.041	0.012	0.041	0.009	0.020	0.004
		0.25	0.049	0.013	0.049	0.009	0.025	0.004
		0.30	0.055	0.014	0.055	0.010	0.029	0.005
	0.2	0.10	0.032	0.010	0.032	0.008	0.010	0.003
		0.15	0.046	0.012	0.046	0.010	0.015	0.003
		0.20	0.057	0.014	0.057	0.011	0.020	0.004
		0.25	0.068	0.015	0.067	0.011	0.025	0.004
		0.30	0.076	0.016	0.076	0.012	0.029	0.005
	0.3	0.10	0.041	0.012	0.040	0.009	0.011	0.003
		0.15	0.057	0.014	0.056	0.011	0.015	0.003
		0.20	0.070	0.015	0.070	0.012	0.020	0.004
		0.25	0.083	0.017	0.082	0.013	0.025	0.005
		0.30	0.094	0.017	0.093	0.013	0.030	0.005
0.3	0.1	0.10	0.039	0.013	0.039	0.010	0.031	0.007
		0.15	0.056	0.014	0.056	0.011	0.044	0.008
		0.20	0.070	0.015	0.070	0.012	0.056	0.009
		0.25	0.083	0.017	0.083	0.012	0.067	0.010
		0.30	0.095	0.018	0.095	0.013	0.078	0.010
	0.2	0.10	0.046	0.013	0.046	0.011	0.031	0.007
		0.15	0.064	0.016	0.065	0.013	0.044	0.009
		0.20	0.079	0.017	0.080	0.014	0.056	0.010
		0.25	0.094	0.018	0.095	0.014	0.067	0.010
		0.30	0.108	0.019	0.108	0.014	0.078	0.011
	0.3	0.10	0.050	0.015	0.050	0.013	0.030	0.008
		0.15	0.071	0.017	0.071	0.014	0.044	0.009
		0.20	0.088	0.017	0.088	0.015	0.056	0.010
		0.25	0.103	0.019	0.104	0.015	0.067	0.011
		0.30	0.117	0.019	0.117	0.015	0.077	0.011

Table B.3: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) across the level of missingness (π), rate of unscheduled visits (λ), and the standard deviation of scheduled visit timings (σ) for three scheduled measurement occasions and sample size (n) 30.

σ	π	λ	**Mean AUC_{OBS}**	**Standard Error**	**Mean AUC_{MLE}**	**Standard Error**	**Mean AUC0**	**Standard Error**
0.1	0.1	0.10	0.023	0.007	0.023	0.004	0.010	0.002
		0.15	0.033	0.008	0.033	0.005	0.016	0.002
		0.20	0.041	0.009	0.041	0.006	0.020	0.003
		0.25	0.048	0.010	0.049	0.007	0.025	0.003
		0.30	0.056	0.011	0.056	0.007	0.030	0.003
	0.2	0.10	0.033	0.008	0.033	0.006	0.011	0.002
		0.15	0.046	0.010	0.046	0.007	0.015	0.002
		0.20	0.058	0.010	0.058	0.008	0.021	0.003
		0.25	0.068	0.012	0.069	0.008	0.025	0.003
		0.30	0.077	0.013	0.078	0.009	0.030	0.003
	0.3	0.10	0.040	0.009	0.040	0.007	0.011	0.002
		0.15	0.056	0.011	0.057	0.008	0.015	0.002
		0.20	0.071	0.012	0.071	0.008	0.021	0.003
		0.25	0.083	0.013	0.084	0.009	0.025	0.003
		0.30	0.094	0.014	0.094	0.009	0.030	0.004
0.3	0.1	0.10	0.040	0.010	0.040	0.007	0.031	0.005
		0.15	0.056	0.012	0.057	0.008	0.045	0.007
		0.20	0.071	0.013	0.071	0.009	0.057	0.007
		0.25	0.084	0.013	0.084	0.009	0.068	0.007
		0.30	0.095	0.014	0.096	0.009	0.079	0.008
	0.2	0.10	0.047	0.011	0.047	0.008	0.031	0.006
		0.15	0.064	0.012	0.065	0.009	0.044	0.006
		0.20	0.082	0.013	0.082	0.010	0.057	0.007
		0.25	0.095	0.014	0.096	0.010	0.068	0.008
		0.30	0.109	0.015	0.109	0.011	0.079	0.008
	0.3	0.10	0.049	0.011	0.050	0.010	0.030	0.006
		0.15	0.071	0.013	0.072	0.011	0.044	0.007
		0.20	0.089	0.014	0.090	0.011	0.057	0.007
		0.25	0.105	0.015	0.105	0.012	0.068	0.008
		0.30	0.118	0.014	0.119	0.011	0.078	0.008

Table B.4: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) across the level of missingness (π), rate of unscheduled visits (λ), and the standard deviation of scheduled visit timings (σ) for five scheduled measurement occasions and sample size (n) 30.

σ	π	λ	**Mean AUC_{OBS}**	**Standard Error**	**Mean AUC_{MLE}**	**Standard Error**	**Mean AUC0**	**Standard Error**
0.1	0.1	0.10	0.023	0.005	0.023	0.003	0.010	0.001
		0.15	0.033	0.006	0.033	0.004	0.015	0.002
		0.20	0.041	0.007	0.041	0.005	0.020	0.002
		0.25	0.049	0.007	0.049	0.005	0.025	0.002
		0.30	0.056	0.008	0.056	0.006	0.030	0.002
	0.2	0.10	0.033	0.006	0.033	0.004	0.010	0.001
		0.15	0.046	0.007	0.046	0.005	0.015	0.002
		0.20	0.058	0.008	0.058	0.006	0.020	0.002
		0.25	0.068	0.008	0.068	0.006	0.025	0.002
		0.30	0.077	0.009	0.077	0.007	0.030	0.003
	0.3	0.10	0.040	0.007	0.040	0.005	0.010	0.001
		0.15	0.056	0.008	0.056	0.006	0.015	0.002
		0.20	0.071	0.009	0.070	0.006	0.020	0.002
		0.25	0.082	0.009	0.083	0.007	0.025	0.002
		0.30	0.094	0.009	0.093	0.007	0.029	0.003
0.3	0.1	0.10	0.040	0.007	0.040	0.005	0.031	0.004
		0.15	0.056	0.008	0.056	0.006	0.044	0.005
		0.20	0.071	0.009	0.071	0.007	0.056	0.005
		0.25	0.083	0.010	0.084	0.007	0.068	0.005
		0.30	0.095	0.010	0.095	0.007	0.078	0.006
	0.2	0.10	0.046	0.008	0.046	0.006	0.031	0.004
		0.15	0.065	0.008	0.065	0.007	0.044	0.005
		0.20	0.082	0.010	0.082	0.007	0.057	0.005
		0.25	0.096	0.010	0.096	0.008	0.068	0.006
		0.30	0.109	0.011	0.109	0.008	0.079	0.006
	0.3	0.10	0.050	0.008	0.050	0.007	0.031	0.004
		0.15	0.071	0.009	0.071	0.008	0.044	0.005
		0.20	0.089	0.010	0.089	0.008	0.057	0.005
		0.25	0.104	0.010	0.105	0.008	0.068	0.006
		0.30	0.118	0.011	0.119	0.008	0.078	0.006

Table B.5: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) across the level of missingness (π), rate of unscheduled visits (λ), and the standard deviation of scheduled visit timings (σ) for three scheduled measurement occasions and sample size (n) 100.

σ	θ	λ	**Mean $\mathbf{AUC_{OBS}}$**	**Standard Error**	**Correct Mean $\mathbf{AUC_{MLE}}$**	**Standard Error**	**Mis-specified Mean $\mathbf{AUC_{MLE}}$**	**Standard Error**
0.05	0.1	0.10	0.018	0.004	0.018	0.003	0.018	0.003
		0.30	0.042	0.007	0.043	0.006	0.045	0.006
	0.5	0.10	0.016	0.003	0.017	0.003	0.019	0.003
		0.30	0.040	0.006	0.042	0.007	0.053	0.006
0.2	0.1	0.10	0.030	0.006	0.034	0.005	0.031	0.004
		0.30	0.075	0.009	0.076	0.008	0.076	0.006
	0.5	0.10	0.025	0.005	0.031	0.004	0.028	0.004
		0.30	0.064	0.008	0.070	0.008	0.071	0.006

Table B.6: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE}) for dependent regular and irregular visit process across the increase in the probability of missingness due to unscheduled visits (θ), rate of unscheduled visits (λ), and the standard deviation of scheduled visit timings (σ) for three scheduled measurement occasions and sample size (n) 100.

σ	θ	λ	**Correct Mean AUC0**	**Standard Error**	**Mis-specified Mean AUC0**	**Standard Error**
0.05	0.1	0.10	0.005	0.001	0.005	0.001
		0.30	0.014	0.001	0.014	0.001
	0.5	0.10	0.005	0.001	0.005	0.001
		0.30	0.014	0.001	0.014	0.001
0.2	0.1	0.10	0.029	0.004	0.021	0.003
		0.30	0.059	0.004	0.056	0.004
	0.5	0.10	0.028	0.003	0.019	0.003
		0.30	0.057	0.004	0.053	0.004

Table B.7: The mean AUCs assuming no missingness (AUC0) for dependent regular and irregular visit process across the increase in the probability of missingness due to unscheduled visits (θ), rate of unscheduled visits (λ), and the standard deviation of scheduled visit timings (σ) for three scheduled measurement occasions and sample size (n) 100.

P_{deact}	P_{act}	λ	λ_{act}	**Mean AUC_{OBS}**	**Standard Error**	**Mean AUC_{MLE}**	**Standard Error**	**Mean AUC0**	**Standard Error**
0.1	0.3	0.10	0.40	0.049	0.007	0.050	0.006	0.026	0.003
			0.60	0.063	0.009	0.065	0.006	0.036	0.003
		0.20	0.40	0.059	0.008	0.059	0.006	0.032	0.003
			0.60	0.070	0.009	0.071	0.007	0.041	0.003
	0.6	0.10	0.40	0.060	0.008	0.061	0.006	0.033	0.003
			0.60	0.077	0.010	0.077	0.007	0.046	0.003
		0.20	0.40	0.067	0.009	0.066	0.006	0.037	0.003
			0.60	0.080	0.010	0.080	0.007	0.049	0.003
0.3	0.3	0.10	0.40	0.042	0.007	0.042	0.005	0.021	0.002
			0.60	0.052	0.008	0.053	0.006	0.028	0.002
		0.20	0.40	0.054	0.008	0.053	0.005	0.028	0.002
			0.60	0.061	0.009	0.062	0.006	0.034	0.003
	0.6	0.10	0.40	0.052	0.008	0.052	0.005	0.027	0.002
			0.60	0.066	0.009	0.066	0.006	0.037	0.003
		0.20	0.40	0.060	0.009	0.060	0.006	0.033	0.003
			0.60	0.071	0.009	0.071	0.007	0.041	0.003

Table B.8: The mean observed AUCs (AUC_{OBS}) and likelihood-based AUCs (AUC_{MLE} and AUC0) for the unobserved process across the probability of the increased rate of unscheduled visits activating and deactivating (P_{act} and P_{deact}), the initial rate of unscheduled visits (λ), the increased rate of unscheduled visits (λ_{act}) for sample size (n) 100.

Number of Scheduled Measurement Occasions (k)	γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
2	0	0.10	0.007	0.004	0.001	0.080
		0.20	0.017	0.006	0.000	0.081
		0.30	0.027	0.008	0.003	0.078
		0.40	0.037	0.008	-0.004	0.077
	0.5	0.10	0.008	0.004	0.073	0.084
		0.20	0.018	0.006	0.118	0.083
		0.30	0.029	0.008	0.145	0.083
		0.40	0.040	0.009	0.166	0.084
	1	0.10	0.011	0.005	0.162	0.095
		0.20	0.023	0.007	0.236	0.095
		0.30	0.035	0.008	0.279	0.101
		0.40	0.045	0.010	0.312	0.104
4	0	0.10	0.004	0.002	0.003	0.075
		0.20	0.012	0.004	-0.001	0.079
		0.30	0.020	0.005	0.003	0.074
		0.40	0.030	0.005	-0.001	0.078
	0.5	0.10	0.005	0.002	0.040	0.079
		0.20	0.013	0.004	0.077	0.082
		0.30	0.023	0.005	0.113	0.080
		0.40	0.032	0.006	0.131	0.082
	1	0.10	0.007	0.003	0.093	0.085
		0.20	0.017	0.005	0.169	0.093
		0.30	0.027	0.006	0.226	0.098
		0.40	0.038	0.007	0.265	0.103

Table B.9: The mean observed AUCs (AUC_{OBS}) and mean bias across the number of scheduled measurement occasions (k), baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100.

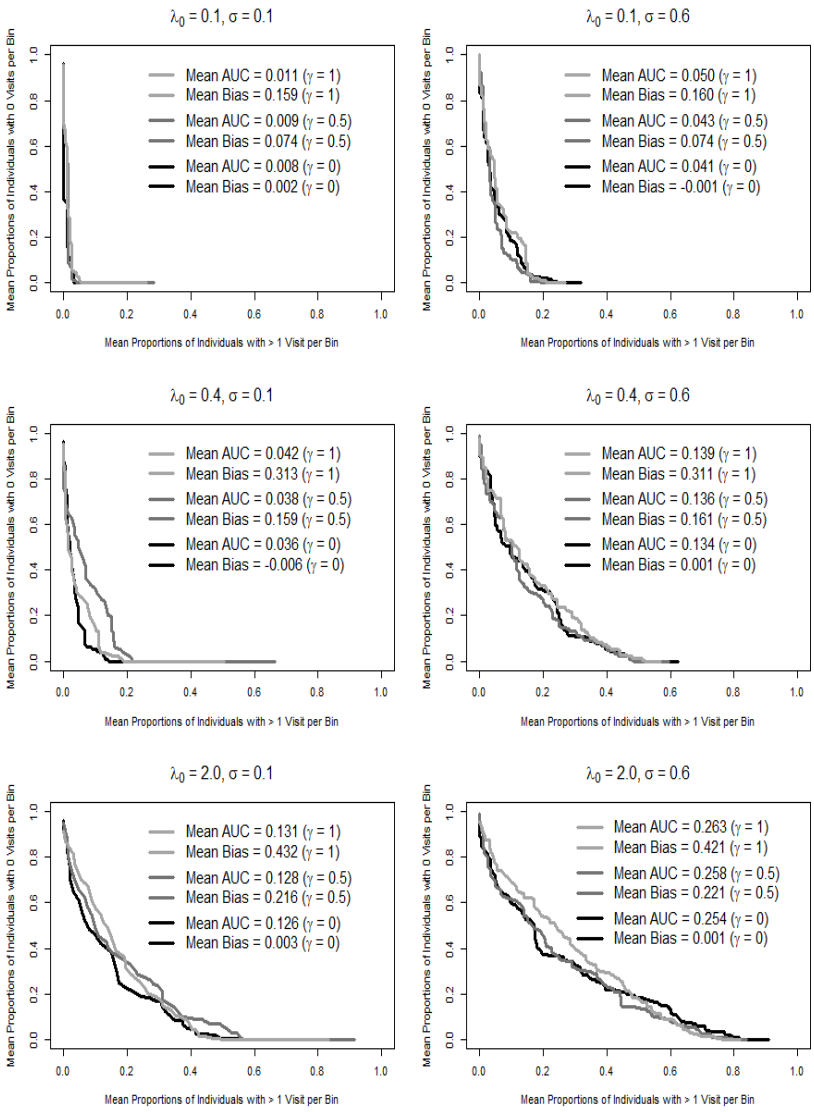

Figure B.1: The mean observed AUCs (AUC_{OBS}) and mean bias across the level of informativeness of the unscheduled visit process (γ) for two scheduled measurement occasions, mean baseline rate of unscheduled visits (λ_0) of $0.1, 0.4, 2.0$, and a standard deviation of scheduled visit timings (σ) of $0.1, 0.6$.

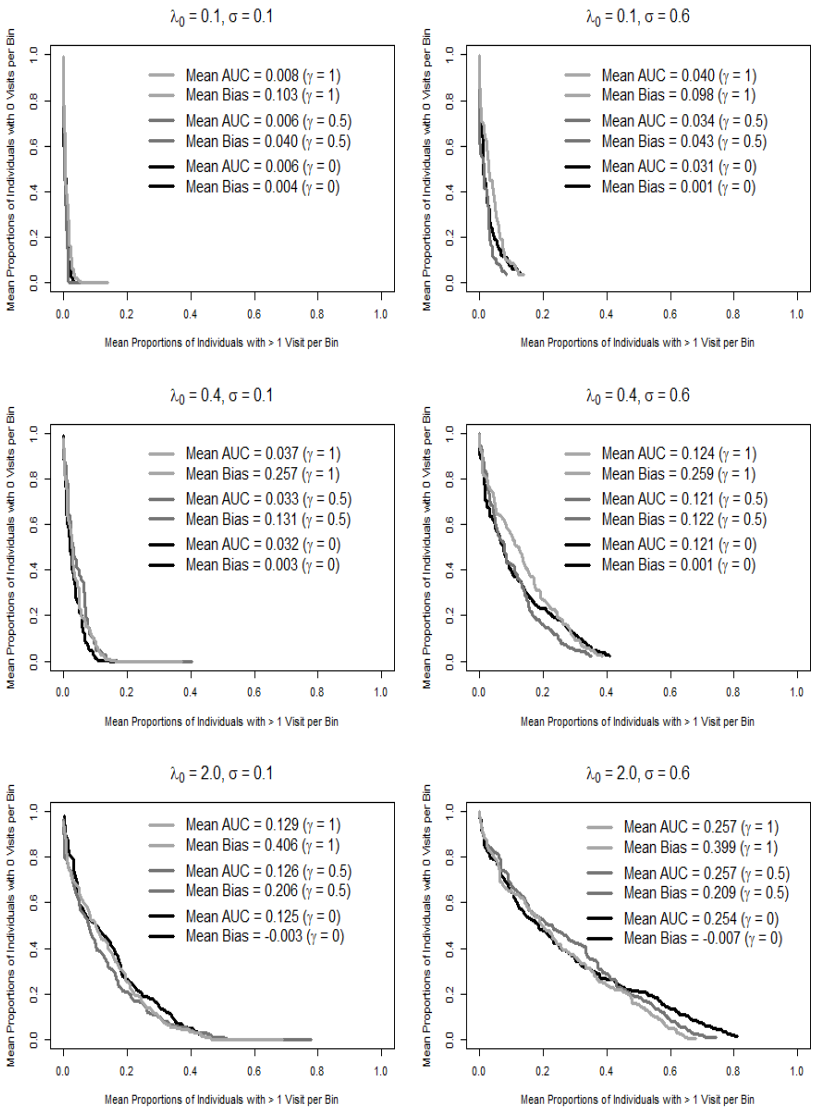

Figure B.2: The mean observed AUCs (AUC_{OBS}) and mean bias across the level of informativeness of the unscheduled visit process (γ) for four scheduled measurement occasions, mean baseline rate of unscheduled visits (λ_0) of $0.1, 0.4, 2.0$, and a standard deviation of scheduled visit timings (σ) of $0.1, 0.6$.

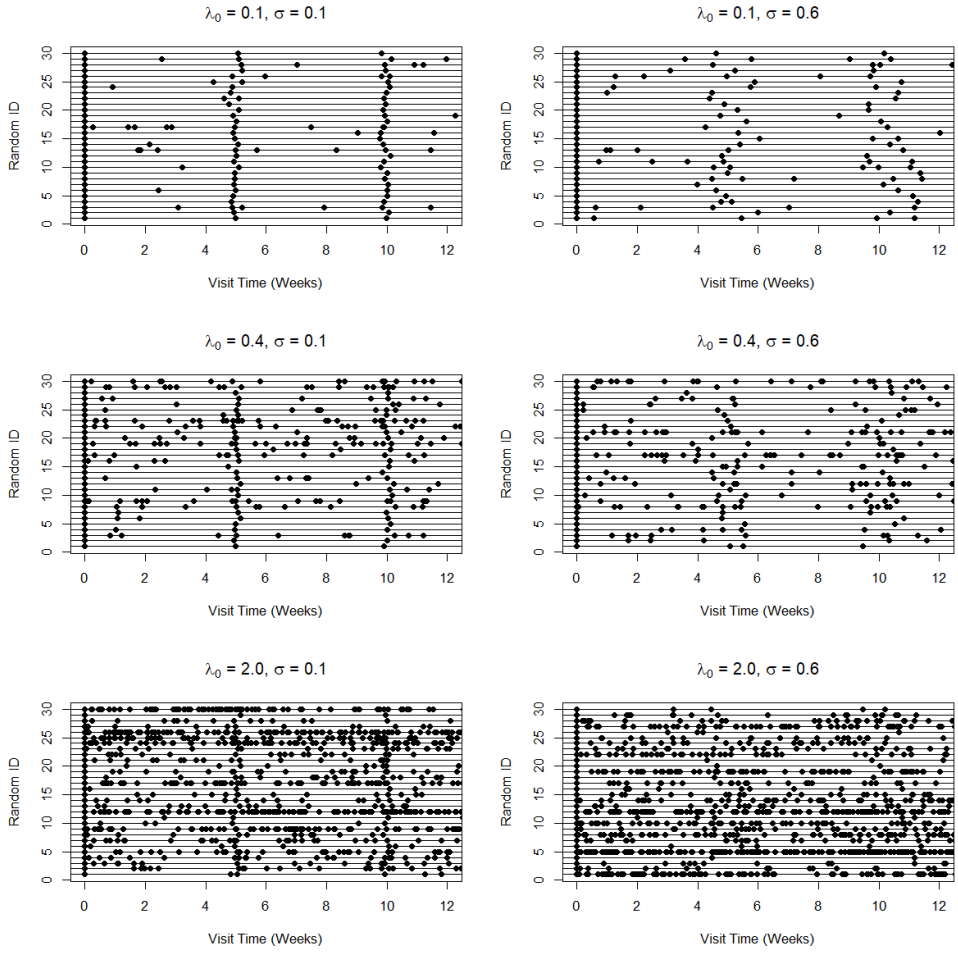

Figure B.3: The visit timings for random subsets of 30 individuals across the level of informativeness of the unscheduled visit process (γ) for two scheduled measurement occasions, mean baseline rate of unscheduled visits (λ_0) of $0.1, 0.4, 2.0$, and a standard deviation of scheduled visit timings (σ) of $0.1, 0.6$.

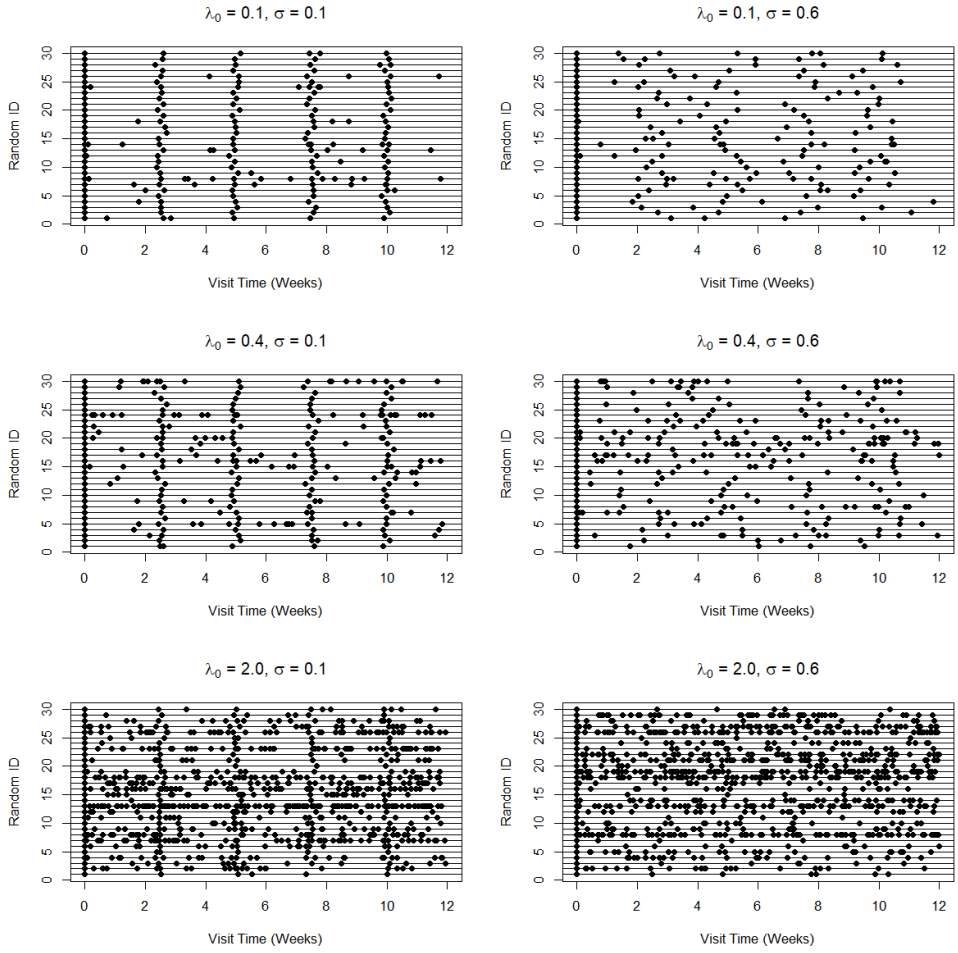

Figure B.4: The visit timings for random subsets of 30 individuals across the level of informativeness of the unscheduled visit process (γ) for four scheduled measurement occasions, mean baseline rate of unscheduled visits (λ_0) of $0.1, 0.4, 2.0$, and a standard deviation of scheduled visit timings (σ) of $0.1, 0.6$.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.003	0.003	0.001	0.079
	0.10	0.008	0.004	0.002	0.086
	0.25	0.023	0.007	-0.004	0.088
	0.40	0.036	0.009	-0.006	0.091
	0.60	0.052	0.011	-0.001	0.093
	1.10	0.085	0.013	0.004	0.096
	1.30	0.096	0.014	0.002	0.095
	2.00	0.126	0.015	0.003	0.098
0.5	0.05	0.004	0.003	0.040	0.084
	0.10	0.009	0.005	0.074	0.092
	0.25	0.024	0.008	0.136	0.097
	0.40	0.038	0.009	0.159	0.104
	0.60	0.054	0.011	0.186	0.104
	1.10	0.087	0.013	0.202	0.105
	1.30	0.098	0.014	0.208	0.105
	2.00	0.128	0.015	0.216	0.101
1	0.05	0.005	0.003	0.095	0.096
	0.10	0.011	0.005	0.159	0.102
	0.25	0.028	0.008	0.264	0.121
	0.40	0.042	0.010	0.313	0.127
	0.60	0.059	0.012	0.350	0.131
	1.10	0.092	0.013	0.394	0.135
	1.30	0.102	0.015	0.405	0.137
	2.00	0.131	0.016	0.432	0.139

Table B.10: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, two scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.1.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.009	0.005	0.001	0.085
	0.10	0.022	0.007	0.003	0.084
	0.25	0.058	0.011	0.002	0.089
	0.40	0.087	0.013	0.001	0.089
	0.60	0.116	0.014	-0.001	0.093
	1.10	0.164	0.016	0.003	0.096
	1.30	0.177	0.016	0.008	0.090
	2.00	0.209	0.016	0.006	0.097
0.5	0.05	0.011	0.005	0.038	0.087
	0.10	0.024	0.007	0.074	0.088
	0.25	0.060	0.012	0.136	0.099
	0.40	0.089	0.014	0.153	0.102
	0.60	0.117	0.014	0.178	0.100
	1.10	0.165	0.016	0.205	0.102
	1.30	0.180	0.016	0.205	0.105
	2.00	0.211	0.016	0.218	0.105
1	0.05	0.014	0.006	0.095	0.096
	0.10	0.029	0.008	0.162	0.105
	0.25	0.066	0.013	0.266	0.125
	0.40	0.093	0.015	0.311	0.129
	0.60	0.122	0.016	0.357	0.137
	1.10	0.168	0.017	0.394	0.137
	1.30	0.181	0.017	0.412	0.132
	2.00	0.214	0.016	0.425	0.141

Table B.11: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, two scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.3.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.018	0.006	0.001	0.081
	0.10	0.041	0.010	-0.001	0.085
	0.25	0.097	0.014	-0.001	0.086
	0.40	0.134	0.015	0.001	0.094
	0.60	0.168	0.016	-0.003	0.088
	1.10	0.216	0.017	0.001	0.094
	1.30	0.229	0.015	0.002	0.098
	2.00	0.254	0.014	0.001	0.097
0.5	0.05	0.020	0.007	0.043	0.084
	0.10	0.043	0.010	0.074	0.090
	0.25	0.099	0.015	0.136	0.097
	0.40	0.136	0.016	0.161	0.096
	0.60	0.170	0.017	0.179	0.098
	1.10	0.218	0.016	0.204	0.106
	1.30	0.230	0.016	0.202	0.107
	2.00	0.258	0.015	0.221	0.110
1	0.05	0.026	0.008	0.092	0.091
	0.10	0.050	0.011	0.160	0.105
	0.25	0.104	0.015	0.265	0.120
	0.40	0.139	0.016	0.311	0.120
	0.60	0.171	0.017	0.344	0.124
	1.10	0.220	0.017	0.395	0.128
	1.30	0.232	0.016	0.409	0.141
	2.00	0.263	0.016	0.421	0.130

Table B.12: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, two scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.6.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.002	0.002	-0.003	0.079
	0.10	0.006	0.003	0.004	0.078
	0.25	0.019	0.005	-0.001	0.081
	0.40	0.032	0.007	0.003	0.085
	0.60	0.049	0.008	-0.003	0.086
	1.10	0.082	0.010	0.001	0.091
	1.30	0.093	0.011	0.003	0.092
	2.00	0.125	0.012	-0.003	0.098
0.5	0.05	0.002	0.002	0.019	0.079
	0.10	0.006	0.003	0.040	0.082
	0.25	0.020	0.005	0.094	0.088
	0.40	0.033	0.007	0.131	0.093
	0.60	0.049	0.008	0.155	0.093
	1.10	0.084	0.010	0.188	0.097
	1.30	0.095	0.012	0.189	0.101
	2.00	0.126	0.012	0.206	0.105
1	0.05	0.004	0.002	0.046	0.084
	0.10	0.008	0.003	0.103	0.094
	0.25	0.023	0.006	0.200	0.102
	0.40	0.037	0.008	0.257	0.117
	0.60	0.054	0.009	0.304	0.120
	1.10	0.087	0.012	0.362	0.126
	1.30	0.097	0.012	0.371	0.132
	2.00	0.129	0.014	0.406	0.132

Table B.13: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, four scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.1.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.006	0.003	-0.005	0.075
	0.10	0.016	0.005	0.002	0.080
	0.25	0.048	0.008	0.001	0.082
	0.40	0.077	0.011	0.004	0.083
	0.60	0.108	0.012	0.003	0.085
	1.10	0.159	0.013	0.001	0.092
	1.30	0.173	0.013	-0.003	0.093
	2.00	0.209	0.013	-0.003	0.093
0.5	0.05	0.007	0.003	0.020	0.078
	0.10	0.018	0.005	0.037	0.082
	0.25	0.050	0.009	0.095	0.087
	0.40	0.078	0.011	0.127	0.093
	0.60	0.109	0.013	0.156	0.098
	1.10	0.160	0.014	0.181	0.101
	1.30	0.174	0.014	0.189	0.102
	2.00	0.209	0.013	0.209	0.104
1	0.05	0.010	0.004	0.044	0.081
	0.10	0.022	0.006	0.093	0.092
	0.25	0.055	0.010	0.199	0.108
	0.40	0.082	0.012	0.250	0.118
	0.60	0.111	0.013	0.303	0.122
	1.10	0.160	0.015	0.368	0.130
	1.30	0.173	0.014	0.375	0.137
	2.00	0.210	0.014	0.406	0.135

Table B.14: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, four scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.3.

γ	λ_0	**Mean AUC_{OBS}**	**Standard Error**	**Mean Bias**	**Standard Error**
0	0.05	0.014	0.004	0.002	0.078
	0.10	0.031	0.007	0.001	0.076
	0.25	0.082	0.011	0.004	0.081
	0.40	0.121	0.012	0.001	0.086
	0.60	0.158	0.013	-0.003	0.087
	1.10	0.211	0.013	-0.001	0.089
	1.30	0.225	0.012	-0.004	0.090
	2.00	0.254	0.011	-0.007	0.094
0.5	0.05	0.016	0.004	0.019	0.076
	0.10	0.034	0.007	0.043	0.081
	0.25	0.084	0.011	0.098	0.089
	0.40	0.121	0.013	0.122	0.090
	0.60	0.158	0.014	0.153	0.094
	1.10	0.212	0.013	0.189	0.100
	1.30	0.225	0.013	0.196	0.105
	2.00	0.257	0.012	0.209	0.109
1	0.05	0.020	0.006	0.052	0.088
	0.10	0.040	0.009	0.098	0.091
	0.25	0.089	0.013	0.196	0.112
	0.40	0.124	0.014	0.259	0.119
	0.60	0.158	0.015	0.295	0.122
	1.10	0.210	0.015	0.366	0.135
	1.30	0.224	0.015	0.378	0.132
	2.00	0.257	0.013	0.399	0.131

Table B.15: The mean observed AUCs (AUC_{OBS}) and mean bias across the mean baseline rate of unscheduled visits (λ_0), and the level of informativeness of the unscheduled visit process (γ) for sample size (n) 100, four scheduled measurement occasions, and the standard deviation of scheduled visit timings (σ) of 0.6.

Appendix C: Worked Example

The Phenobarb dataset from the "MEMSS" package in R [63] is used to illustrate the estimation of the mean proportions of individuals with 0, 1, and >1 visits per bin, and the area under the curve (AUC) of the mean proportions of individuals with 0 vs. >1 visits per bin plot. Clinical pharmacokinetic data were collected from 59 preterm infants who received phenobarbital in order to prevent seizures. Blood draws were taken to determine serum concentration of phenobarbital, with the timing and number of draws varying among infants. For the purposes of illustration, this analysis focuses not on pharmacokinetics but on the mean serum concentration of phenobarbital over time.

This dataset contains some rows corresponding to times at which phenobarbital given and others at which a blood draw was taken to determine serum phenobarbital concentration. Analysis shows that once previous serum concentration is accounted for, dose is uninformative about the timing of blood draws, and hence we remove rows at which concentration was not measured. Moreover, estimating the mean proportions of individuals with 0, 1, and >1 visits per bin and the AUC requires the variable identifying subjects to be numeric rather than a factor, so we create a new numeric "ID" variable. The AUC is estimated using the "zoo" package [64].

The function "meanproportions" returns the estimated mean proportions of individuals with 0, 1, and >1 visits per bin and the AUC.

```
library(MEMSS)
library(zoo)
```

```
data("Phenobarb")
PC <- Phenobarb[is.finite(Phenobarb$conc),]
PC$ID = as.numeric(PC$Subject)
```

```
meanproportions<-function(data, times, ID, maxnumbins)
{
  meang=c(0)
  mean0=c(0)
  mean11=c(0)
  times = as.numeric(times)
  ID = as.numeric(ID)

  #Estimating the Mean proportions of individuals
  #with 0, 1, and > 1 Visits per bin and AUC
```

```
  for(j in 1:maxnumbins)
  {
    cutpoints<-seq(0,max(times), (max(times)/j))
    percg<-c(0)
    perc1<-c(0)
    perc0<-c(0)

    data$bin<-cut(times, cutpoints, include.lowest=TRUE)
        obsbin1<-table(ID, data$bin)
    gbin1<-apply(obsbin1>1,2,sum)
    obin1<-apply(obsbin1==1,2,sum)
    nonebin1<-apply(obsbin1==0,2,sum)

    for(i in 1:j)
    {
      percg[i]<-gbin1[i]/length(unique(ID))
      perc1[i]<-obin1[i]/length(unique(ID))
      perc0[i]<-nonebin1[i]/length(unique(ID))
    }

    meang[j]=mean(percg)
    mean0[j]=mean(perc0)
    mean11[j] = mean(perc1)
  }
   #Estimating the area under the curve using the "zoo" R package
  AUCobs <- round(sum(diff(meang[order(meang)])*
  rollmean(mean0[order(meang)],2)), 3)

  results = list(AUCobs, meang, mean0, mean11)
  names(results) <- c("AUCobs", "meang", "mean0", "mean11")
  list2env(results, envir = .GlobalEnv)
}
```

The visit timings for a random subset of 30 infants, the estimated mean proportions of individuals with 0, 1, and >1 visits per bin, and the AUC are plotted.

```
maxnumbins = 100
meanproportions(PC, PC$time, PC$ID, maxnumbins)
```

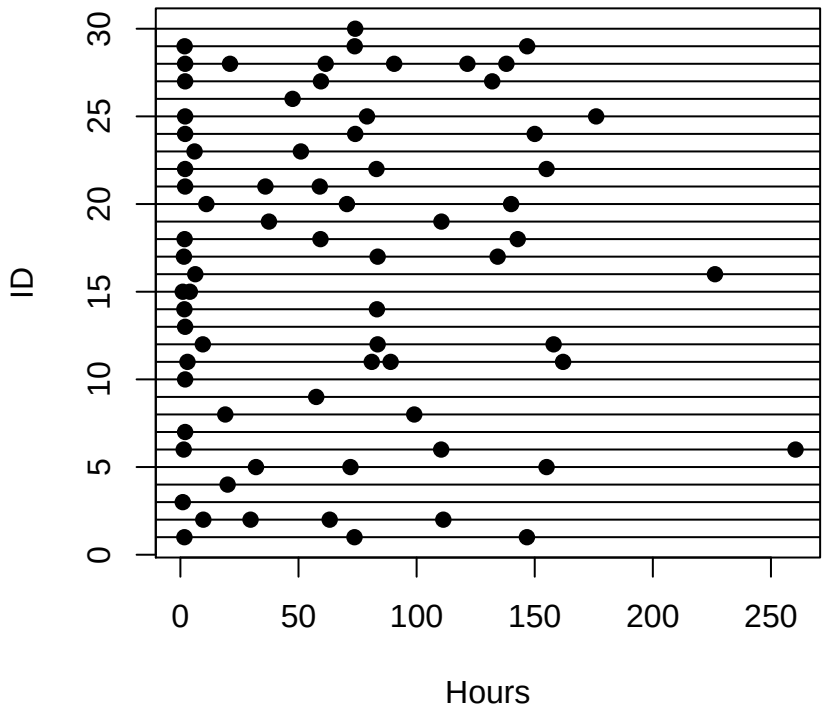

Figure C.1: The visit timings (hours) for a random subset of 30 infants from the Phenobarb dataset.

```
sampleids = sample(unique(PC$ID), size = 30)

rs = PC[PC$ID %in% sampleids,]
rs$ID1 = c(0)
for(i in 1:30)
{
  for(j in 1:length(rs$ID1))
  {
    if(rs$ID[j] == sampleids[i]){rs$ID1[j] = i}
  }
}

plot(rs$time, rs$ID1, pch=19, lty=2, xlim=c(0,max(rs$time)), xlab="Hours", ylab="ID")
abline(h = c(1:30))
```

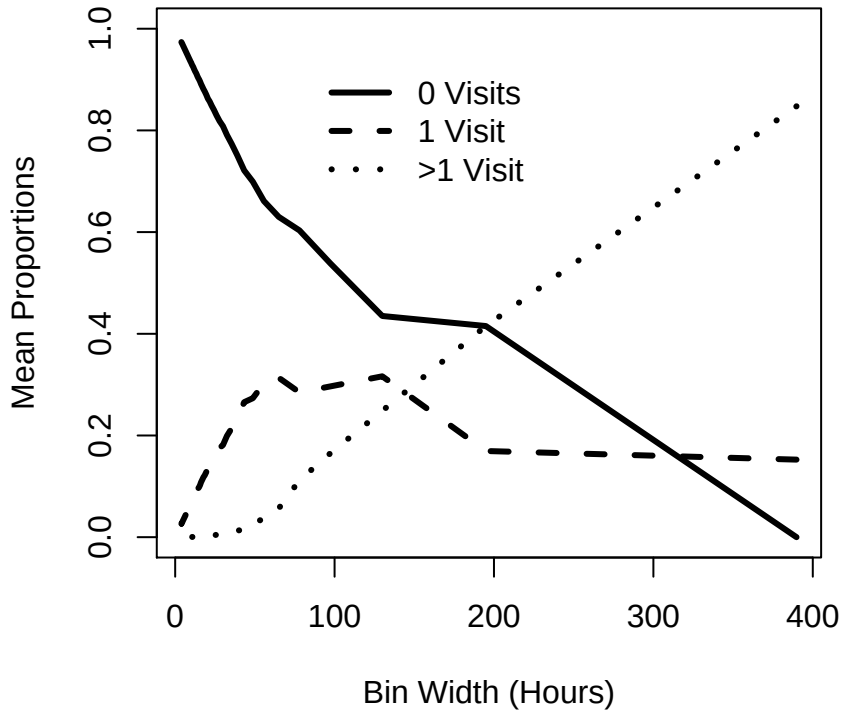

Figure C.2: The mean proportions of individuals with 0, 1, and >1 visits per bin across bin width.

```
  plot(x = max(PC$time)/c(1:maxnumbins), y=mean0, type="l", lwd=3,
 ylim=c(0,1), xlab="Bin Width (Hours)", ylab="Mean Proportions")
  lines(x = max(PC$time)/c(1:maxnumbins), y=mean11, lwd=3,lty=2)
  lines(x = max(PC$time)/c(1:maxnumbins), y=meang, lwd=3,lty=3)
  legend(80, 0.95, legend=c("0 Visits",
"1 Visit",">1 Visit"), lwd=c(3,3,3),lty=c(1,2,3), bty="n")
```

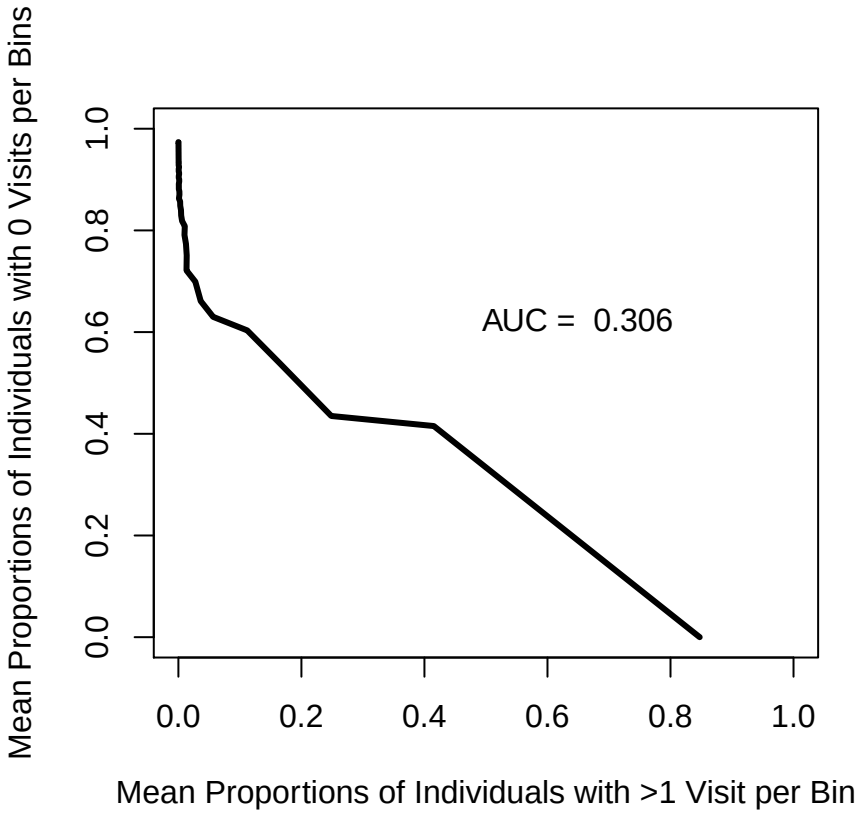

Figure C.3: The mean proportions of individuals with 0 vs. >1 visits per bin and the resulting area under the curve (AUC).

```
  plot(x = meang, y=mean0, type="l", lwd=3, xlim=c(0,1),
ylim=c(0,1), xlab="Mean Proportions of Individuals with >1 Visit per Bin",
ylab="Mean Proportions of Individuals with 0 Visits per Bins")
  legend(0.4, 0.7, legend=c(paste("AUC = ", AUCobs)), bty="n")
```

Bibliography

[1] P. J. Diggle, P. J. Heagerty, K. Y. Liang, and S. L. Zeger. *Analysis of longitudinal data*. Oxford Statistical Science Serires, second edition edition, 2002.

[2] T. R. Dawber, G. F. Meadors, and F. E. Moore. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health*, 41(3):279–281, 1951.

[3] W. J. Rogan, K. N. Dietrich, J. H. Ware, D. W. Dockery, M. Salganik, J. Radcliffe, R. L. Jones, N. B. Ragan, J. J. Chisolm, and G. G. Rhoads. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. *N. Engl. J. Med.*, 344(19):1421–1426, 2001.

[4] S. R. Sukumar, R. Natarajan, and R. K. Ferrell. Quality of Big Data in health care. *Int J Health Care Qual Assur*, 28(6):621–634, 2015.

[5] T. Gomes, D. N. Juurlink, I. A. Dhalla, A. Mailis-Gagnon, J. M. Paterson, and M. M. Mamdani. Trends in opioid use and dosing among socio-economically disadvantaged patients. *Open Med*, 5(1):13–22, 2011.

[6] R. Sutradhar, A. Lokku, and L. Barbera. Cancer survivorship and opioid prescribing rates: A population-based matched cohort study among individuals with and without a history of cancer. *Cancer*, 123(21):4286–4293, 2017.

[7] L. SH. Lim, E. M. Pullenayegum, L. Lim, D. Gladman, B. Feldman, and E. Silverman. From childhood to adulthood: The trajectory of damage in patients with juvenile-onset systemic lupus erythematosus. *Arthritis Care & Research*, 69(11):1627–1635, 2017.

[8] S. Carsley, C. M. Borkhoff, J. L. Maguire, C. S. Birken, M. Khovratovich, B. McCrindle, C. Macarthur, and P. C. Parkin. Cohort profile: The applied research group for kids (target kids!). *International Journal of Epidemiology*, 44(3):776–788, 2014.

[9] K. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[10] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25, 1993.

[11] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 221–233, Berkeley, Calif., 1967. University of California Press.

[12] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

[13] P. J. Heagerty and B. F. Kurland. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88(4):973–985, 2001.

[14] J. Ferron, R. Dailey, and Q. Yi. Effects of Misspecifying the First-Level Error Structure in Two-Level Models of Change. *Multivariate Behav Res*, 37(3):379–403, 2002.

[15] E. M. Pullenayegum and L. SH. Lim. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statistical Methods in Medical Research*, 25(6):2992–3014, 2016.

[16] R. J. Cook and J. F. Lawless. Analysis of repeated events. *Statistical Methods in Medical Research*, 11(2):141–166, 2002.

[17] Z. Guo, T. M. Gill, and H. G. Allore. Modeling repeated time-to-event health conditions with discontinuous risk intervals. An example of a longitudinal study of functional disability among older persons. *Methods Inf Med*, 47(2):107–116, 2008.

[18] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

[19] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[20] H. Lin, D. O. Scharfstein, and R. A. Rosenheck. Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813, 2004.

[21] P. Bůžková and T. Lumley. Semiparametric modeling of repeated measurements under outcome-dependent follow-up. *Statistics in Medicine*, 28(6):987–1003, 2009.

[22] Y. Liang, W. Lu, and Z. Ying. Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2):377–384, 2009.

[23] L. Sun, X. Mu, Z. Sun, and X. Tong. Semiparametric analysis of longitudinal data with informative observation times. *Acta Mathematicae Applicatae Sinica, English Series*, 27(1):29–42, 2011.

[24] X. Song, X. Mu, and L. Sun. Regression analysis of longitudinal data with time-dependent covariates and informative observation times. *Scandinavian Journal of Statistics*, 39(2):248–258, 2012.

[25] L. Sun, X. Song, J. Zhou, and L. Liu. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498):688–700, 2012.

[26] W. Su and H. Jiang. Semiparametric analysis of longitudinal data with informative observation times and censoring times. *Journal of Applied Statistics*, 45(11):1978–1993, 2018.

[27] D. Machin, T. M. Farley, B. Busca, M. J. Campbell, and C. d'Arcangues. Assessing changes in vaginal bleeding patterns in contracepting women. *Contraception*, 38(2):165–179, 1988.

[28] D. Scharfstein, A. McDermott, I. Diaz, M. Carone, N. Lunardon, and I. Turkoz. Global sensitivity analysis for repeated measures studies with informative drop-out: A semi-parametric approach. *Biometrics*, 74(1):207–219, 2018.

[29] J. J. Deeks, J. PT. Higgins, and D. G. Altman. *Cochrane Handbook for Systematic Reviews of Interventions*. 2008.

[30] J. P. Higgins and S. G. Thompson. Quantifying heterogeneity in a meta-analysis. *Stat Med*, 21(11):1539–1558, 2002.

[31] D. A. Belsley, E. Kuh, and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. 1980.

[32] T. A. Craney and J. G. Surles. Model-dependent variance inflation factor cutoff values. *Quality Engineering*, 14(3):391–403, 2002.

[33] R. Little and D. Rubin. *Statistical Analysis with Missing Data, Second Edition*. Wiley, 2014.

[34] A. Lokku, L. SH. Lim, C. S. Birken, and E. M. Pullenayegum. Summarizing the extent of visit irregularity in longitudinal data. *BMC Medical Research Methodology*, 20:135, 2020.

[35] P. Bůžková, E. R. Brown, and G. C. John-Stewart. Longitudinal data analysis for generalized linear models under participant-driven informative follow-up: an application in maternal health epidemiology. *Am. J. Epidemiol.*, 171(2):189–197, 2010.

[36] D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[37] J. P. Vandenbroucke, E. Elm, D. G. Altman, P. C. Gøtzsche, C. D. Mulrow, S. J. Pocock, C. Poole, J. J. Schlesselman, M. Egger, and for the STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *Annals of Internal Medicine*, 147(8):163–194, 2007.

[38] K. Schulz, D. Altman, and D. Moher. Consort 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and Pharmacotherapeutics*, 1(2):100–107, 2010.

[39] D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, 1987.

[40] J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

[41] D. Farzanfar, A. Abumuamar, J. Kim, E. Sirotich, Y. Wang, and E. M. Pullenayegum. Longitudinal studies that use data collected as part of usual care risk reporting biased results: a systematic review. *BMC Medical Research Methodology*, 17(1):133, 2017.

[42] R. Pivovarov, D. J. Albers, G. Hripcsak, J. L. Sepulveda, and N. Elhadad. Temporal trends of hemoglobin A1c testing. *J Am Med Inform Assoc*, 21(6):1038–1044, 2014.

[43] E. M. Pullenayegum. *IrregLong: Analysis of Longitudinal Data with Irregular Observation Times*, 2019. R package version 0.1.0.

[44] F. E. Matthews, M. Chatfield, C. Freeman, C. McCracken, and C. Brayne. Attrition and bias in the MRC cognitive function and ageing study: an epidemiological investigation. *BMC Public Health*, 4:12, 2004.

[45] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[46] D. D. Gladman, C. H. Goldsmith, M. B. Urowitz, P. Bacon, C. Bombardier, D. Isenberg, K. Kalunian, M. H. Liang, P. Maddison, and O. Nived. Crosscultural validation and reliability of 3 disease activity indices in systemic lupus erythematosus. *J. Rheumatol.*, 19(4):608–611, 1992.

[47] D. D. Gladman, D. Ibanez, and M. B. Urowitz. Systemic lupus erythematosus disease activity index 2000. *J. Rheumatol.*, 29(2):288–291, 2002.

[48] Terry M Therneau. *A Package for Survival Analysis in S*, 2015. version 2.38.

[49] D. M. Farewell, C. Huang, and V. Didelez. Ignorability for general longitudinal data. *Biometrika*, 104(2):317–326, 2017.

[50] S. R. Lipsitz, G. M. Fitzmaurice, J. G. Ibrahim, R. Gelber, and S. Lipshultz. Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58(3):621–630, 2002.

[51] E. M. Pullenayegum. Multiple outputation for the analysis of longitudinal data subject to irregular observation. *Statistics in Medicine*, 35(11):1800–1818, 2016.

[52] Within-cluster resampling. *Biometrika*, 88(4):1121–1134, 2001.

[53] J. M. Neuhaus, C. E. McCulloch, and R. D. Boylan. Analysis of longitudinal data from outcome-dependent visit processes: Failure of proposed methods in realistic settings and potential improvements. *Statistics in Medicine*, 37(29):4457–4471, 2018.

[54] B. Gaynes, A. Rush, M. Trivedi, S. Wisniewski, D. Spencer, and M. Fava. The star*d study: Treating depression in the real world. *Cleveland Clinic journal of medicine*, 75:57–66, 2008.

[55] M. H. Trivedi, J. A. Rush, S. R. Wisniewski, A. A. Nierenberg, D. Warden, L. Ritz, G. Norquist, R. H. Howland, B. Lebowitz, P. J. McGrath, K. Shores-Wilson, M. M. Biggs, G. K. Balasubramani, M. Fava, and STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in star*d: Implications for clinical practice. *American Journal of Psychiatry*, 163(1):28–40, 2006.

[56] A. J. Rush, M. H. Trivedi, S. R. Wisniewski, A. A. Nierenberg, J. W. Stewart, D. Warden, G. Niederehe, M. E. Thase, P. W. Lavori, B. D. Lebowitz, P. J. McGrath, J. F. Rosenbaum, H. A. Sackeim, D. J. Kupfer, J. Luther, and M. Fava. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A star*d report. *American Journal of Psychiatry*, 163(11):1905–1917, 2006.

[57] Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011.

[58] U. Halekoh and S. Hojsgaard. *The R Package geepack for Generalized Estimating Equations*, 2006. R package version 1.3-1.

[59] Shaun Seaman, Ian White, Andrew Copas, and Leah Li. Combining multiple imputation and inverse-probability weighting. *Biometrics*, 68:129–37, 11 2011.

[60] M. M. Joffe. Administrative and artificial censoring in censored regression models. *Statistics in Medicine*, 20(15):2287–2304, 2001.

[61] A. Stromberg. Why write statistical software? the case of robust statistical methods. *Journal of Statistical Software, Articles*, 10(5):1–8, 2004.

[62] Weisberg S. and Fox J. *An R Companion to Applied Regression*. Thousand Oaks: Sage, 2 edition, 2011.

[63] D. Bates, M. Maechler, and B. Bolker. *MEMSS: Data Sets from Mixed-Effects Models in S*, 2019. R package version 0.9-3.

[64] A. Zeileis and G. Grothendieck. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, 14(6):1–27, 2005.