

Medication Adherence: Definitions, Calculations, and Statistical Modeling Strategies

By

Joshua Joseph DeClercq

Thesis

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

August 10, 2018

Nashville, Tennessee

Approved:

Leena Choi, Ph.D.

Robert Greevy, Ph.D.

## ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to Dr. Leena Choi for all of her help and support in writing this thesis. She provided guidance when I ran into struggles and encouragement when I needed it most. I especially appreciate her dedication to helping me reach my desired deadline. Not only has she been a great thesis advisor, but she has also been a fantastic mentor. She has been integral in my development as a statistician, and I am happy to have the privilege to continue working with her.

Special thanks are due to my thesis committee member Dr. Robert Greevy, for believing in me and the work presented here. Furthermore, his Intro to Biostatistics class provided a very strong foundation for everything that has come since. My entire experience in the Biostatistics Graduate Program has been filled with amazing courses, teachers and fellow students. I could not have asked for a more positive experience over the past two years. Special thanks to Dr. Jeffrey Blume for convincing me that Vanderbilt was the right choice for me. Within my cohort, I have had the pleasure to learn alongside some of the brightest and kindest people I have ever met. I may have been able to get through this program without them, but it would have been nowhere near as much fun.

Most importantly, I would like to thank my family. My wife Courtney gave me so much strength and encouragement. She always believed in me, and kept me going when things got rough. My daughter Eleanor is an unending source of inspiration. And finally, I would like to thank my dog Bowie. He's a good boy.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS . . . . .	ii
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	vii
1 Introduction . . . . .	1
1.1 Background . . . . .	1
1.2 Commonly used measures . . . . .	3
1.2.1 Primary adherence . . . . .	3
1.2.2 Secondary adherence . . . . .	3
1.2.3 Gap-based adherence measures . . . . .	5
1.2.4 Measures of persistence . . . . .	6
1.2.5 Trajectory measures . . . . .	7
1.3 Efforts towards standardization . . . . .	8
1.3.1 AdhereR package . . . . .	9
1.4 Further considerations and limitations . . . . .	10
1.5 Unification of adherence measures . . . . .	12
1.5.1 Denominator calculation . . . . .	13
1.5.2 Numerator calculation . . . . .	15
1.5.3 Drug switching . . . . .	15
1.5.4 Multiple concurrent medications . . . . .	18
1.5.5 Additional considerations . . . . .	19
1.5.6 Reporting MAP . . . . .	19
1.6 Classification of persistence measures . . . . .	20

2	Statistical Methods . . . . .	23
2.1	Introduction . . . . .	23
2.2	Logistic regression . . . . .	24
2.3	Ordinal logistic regression . . . . .	25
2.4	Negative binomial regression . . . . .	27
2.5	Longitudinal regression modeling . . . . .	29
2.5.1	GEE for non-permissible gaps . . . . .	30
2.5.2	GEE for continuous gap and surplus time . . . . .	31
3	Case study . . . . .	33
3.1	Data . . . . .	33
3.1.1	Patient demographics . . . . .	33
3.1.2	PDC calculations by different methods . . . . .	35
3.1.3	Results of PDC calculations . . . . .	38
3.1.4	Gap-based outcomes . . . . .	39
3.2	Sensitivity analysis . . . . .	42
3.2.1	Logistic regression . . . . .	43
3.2.2	Ordinal regression . . . . .	44
3.2.3	Negative binomial regression . . . . .	46
3.2.4	GEE regression . . . . .	47
3.2.5	Model comparison . . . . .	48
3.3	Model checking and comparison . . . . .	49
3.3.1	Predicted PDC . . . . .	50
3.3.2	Predicted PDC $\geq 80\%$ . . . . .	50
3.3.3	Predicted gap days . . . . .	52
3.3.4	Predicted gap and surplus days . . . . .	52
3.3.5	Predicted median PDC . . . . .	52

4	Simulation study . . . . .	54
4.1	Model specifications . . . . .	54
4.2	Simulation Results . . . . .	57
4.2.1	Power and type I error . . . . .	58
4.2.2	Predicted PDC . . . . .	58
4.2.3	Predicted PDC $\geq 80\%$ . . . . .	61
4.2.4	Predicted gap days . . . . .	61
4.2.5	Predicted median PDC . . . . .	62
4.2.6	GEE model performance . . . . .	62
4.3	Ordinal model simulations . . . . .	64
4.3.1	Under similar adherence conditions . . . . .	64
4.3.2	Under less favorable adherence conditions . . . . .	65
5	Discussion . . . . .	68
	BIBLIOGRAPHY . . . . .	70

## LIST OF TABLES

Table	Page
3.1 Patient demographics from the MS study . . . . .	34
3.2 Days supply administered at each fill. . . . .	35
3.3 Summary statistics of PDC calculations . . . . .	37
3.4 Sensitivity analysis - logistic model . . . . .	44
3.5 Sensitivity analysis - ordinal model . . . . .	45
3.6 Sensitivity analysis - negative binomial model . . . . .	46
3.7 Sensitivity analysis - GEE model . . . . .	47
3.8 Method comparison . . . . .	48
3.9 Model checking and comparison - predicted mean PDC . . . . .	51
3.10 Model checking and comparison - probability of $PDC \geq 80\%$ . . . . .	51
3.11 Model checking and comparison - predicted gap days . . . . .	52
3.12 Model checking and comparison - predicted gap and surplus days . . . . .	53
3.13 Model checking and comparison - predicted median PDC . . . . .	53
4.1 Simulation Results - Average Cohort Statistics . . . . .	57
4.2 Simulation Results - Power and Type I Error . . . . .	58
4.3 Simulation Results - Predicted Mean PDC . . . . .	60
4.4 Simulation Results - Probability of $PDC \geq 80\%$ . . . . .	61
4.5 Simulation Results - Estimation of Gap Days . . . . .	62
4.6 Simulation Results - Estimation of Median PDC . . . . .	63
4.7 Simulation Results - GEE Model Performance . . . . .	63
4.8 Ordinal Model Simulation Results - Power and Type I Error . . . . .	65
4.9 Ordinal Model Simulation Results - Predicted Mean PDC . . . . .	65
4.10 Low Adherence Simulation Results - Power and Type I Error . . . . .	66
4.11 Low Adherence Simulation Results - Predicted Mean PDC . . . . .	67

## LIST OF FIGURES

Figure	Page
1.1 Calculation of PDC based on different end date rules. . . . .	16
2.1 Histogram of PDC . . . . .	24
3.1 PDC comparison based on different end date rules . . . . .	37
3.2 Cumulative number of enrolled patients by total number of fills . . . . .	39
3.3 Total number of cumulative gap days per patient . . . . .	40
3.4 Proportion of patients with non-permissible gap by various thresholds for days without medication . . . . .	41
3.5 Proportion of patients by various thresholds to dichotomize permissible ver- sus non-permissible based on interval PDC . . . . .	41
3.6 Timing of fills at each interval . . . . .	42
4.1 Cumulative number of enrolled patients by treatment status . . . . .	56

# Chapter 1

## Introduction

### 1.1 Background

In the United States, more than half of all adults are on at least one prescription drug, and between 2000 and 2012, that number increased from 51% to 59% [1]. Patient adherence to medication is defined as the extent to which a patient takes prescribed medications according to the dosage and frequency recommended by the provider [2]. Medication non-adherence is a widespread problem and has been associated with worse health outcomes, more hospitalizations and increased healthcare costs [3]. Uniform measurements, calculations and operational definitions are not consistently implemented in the area of adherence research, and some adherence-related publications do not carefully define their terminology or methodology, leading to much confusion about the chosen metrics [4]. Without a uniform conceptual framework, medication adherence research is not generalizable. Effective interventions are required to improve adherence, and are predicated on a standardized definition of adherence, a transparent method of calculation, and robust statistical modeling methods.

The act of adhering to a medication regimen is comprised of a series of health behaviors. Outside of being physically present while observing the patient take the medication, adherence to the prescribed medication must be measured by proxy. Some methods used to measure medication adherence include: blood tests, assessment of the patient's clinical response, self-reporting, electronic medication reminders, digital pills, pill counts, patient activation measures (PAMs) and group trajectory model measures [5]. The latter two are based on the act of filling a medication prescription, which can be measured by using electronic databases such as pharmacy dispensing databases or electronic health records (EHRs) [4]. Claims-based measures are used in health plan quality ratings, in identifying



non-adherent patients for targeted interventions by health systems, in research investigating the impact of medication non-adherence on clinical outcomes, and in predicting health care costs and utilization [6]. While attractive because of the reduced costs compared to the other methods, the use of EHRs as a proxy measure for adherence has its own set of inherent problems. The primary problem is that having a prescription filled is not the same as taking the medication in the correct fashion.

The use of claims records for tracking prescription data to measure adherence is often categorized into two measures: medication adherence and persistence. Adherence is the degree to which the patient conforms to the medication use recommendations specified by the prescriber, and can be further categorized into two sub-classifications: primary and secondary adherence. Primary adherence is a measure of whether or not the patient received the first prescription, whereas secondary adherence is an ongoing measure of whether or not the patient received dispensings over the course of a defined period of observation. Persistence to medication is defined as the act of continuing the treatment for the prescribed duration. Contained within medication persistence is the implication that the patient has exhibited at least primary adherence to their prescription regimen. Persistence can be broken down into two categories, early-stage and later-stage persistence [4]. The converse of the discussed terms is simply denoted with the prefix “non-”, such as primary non-adherence or later-stage non-persistence.

The objective of adherence research should be to identify a measurement or a set of measurements that will adequately assess medication adherence and persistence to the maximum degree allowable by the nature of using EHRs. In this thesis, we present a summary of common measures and their variants, discuss their strengths and limitations, and suggest viable alternatives.

## 1.2 Commonly used measures

### 1.2.1 Primary adherence

In 2011, the Pharmacy Quality Alliance (PQA), convened an expert panel to develop a quality measure for primary medication non-adherence (PMN). According to the PQA, “PMN occurs when a new medication is prescribed for a patient, but the patient does not obtain the medication, or appropriate alternative, within an acceptable period of time after it was prescribed” [7]. Adams et al. put forth the definition of an acceptable period of time as being within 30 days of the prescribing event. Additionally, if a prescription is reversed and not collected by the patient, it is not considered a dispensing event. Because patients can either be only primary adherent or non-adherent, measurement at the individual level is a dichotomous outcome, based on whether or not the prescription was filled within the prespecified time frame. At the group or study level, the calculation of PMN can be expressed as a ratio. The denominator for this calculation consists of the number of new prescriptions for a drug therapy during the measurement period, and the numerator is the number of prescribing transactions where there is no record of medication dispensation [7].

There are many potential limitations to the PMN measurement. Notably, it is only compatible with electronic prescription and transaction data. Changes in pharmacy benefits may confound the calculations. Additionally, it does not account for instances where the patient is given a medication sample at the time of prescription, or that the prescription was filled at a different pharmacy rather than the one at which it was prescribed.

### 1.2.2 Secondary adherence

The Medication Possession Ratio (MPR) is a commonly used method for claims-based adherence measurement. However, despite its widespread use, its operational definition is highly variable across its practitioners. In general, the calculation is derived as the number of days within a prespecified time frame for which a patient has the prescribed medication

(e.g., the days supply of medication) divided by the number of days in the study period interval, and is often reported as a percentage rather than a ratio. An alternative method is MPR modified (MPRm), which is defined as the days supply of medication dispensed during the specified observation period excluding the last refill, divided by the number of days between the first and last dispensing [4].

Because it is possible for patients to obtain a larger days supply than the defined duration of the study period, some calculations of MPR are allowed to exceed 100%, while others are capped at 100% adherence. By not capping the measure at 100%, this will overestimate true rate of medication adherence when calculated at the population level. Furthermore, MPR is often calculated across a medication class (e.g., all statin drugs), and therefore a switch and overlap of medications in the same class during the study period will inflate the measurement. Some promote the ability of MPR to go above 100% as strength, citing that a MPR greater than 100% is a measure of over-adherence [8]. The MPR does a poor job of measuring adherence when calculated over a short time, and often in its calculation, those with primary medication non-adherence are excluded [9]. In order to be accurate, using MPR to measure adherence requires that patients obtain their medications from a closed pharmacy system such that all pharmacy records are available for the duration the patient is on study [5]. Lastly, MPR is often reported as a categorical measure, usually an indicator if the patient has an  $MPR < 80\%$  or  $\geq 80\%$ .

Proportion of Days Covered (PDC) is a newer measure than MPR. In 2012, the PQA declared that it was the preferred method of measuring medication adherence [9]. Although some variations in the calculation do exist, its operational definition is more consistent than that of MPR. Like MPR, PDC is based on the fill dates and days supply for each prescription fill. The PDC is expressed as a ratio of the number of days covered by the prescription fills over the duration of the observation window, defined as the date of the first fill to the end of the study period. Rather than a summation of the days supply, prescription fills are entered as time arrays. If one time array overlaps with another (i.e., the patient refills the

prescription prior to using the current supply of medication), the new time array is shifted to begin once the older prescription has been used up. As opposed to MPR, this method does not result in values greater than 100% [9].

The PDC provides a more conservative estimate of the adherence rate compared to MPR in situations when the patient has switches of medications within a class or concurrently uses more than one drug in a class. In the former case, the arrays are shifted as described, whereas under the latter scenario, the arrays are not shifted. This method reflects whether the patient had at least one of the concurrent medications available on a particular day. Another strength listed by the PQA is that adjustment for inpatient hospital stays does not significantly alter the population estimate for adherence, even within a population that is prone to frequent inpatient visits [9].

When discussing the calculation of adherence, especially with PDC, the details of important considerations are often omitted. A typical definition encountered in the literature is the ratio of the number of days covered by the prescription fills over the duration of the study period. The face-value interpretation of this definition will lead to something more akin to MPR being what is calculated. A clearer description of the calculation that captures the intent of the metric - though not all of the nuance - would be something like the ratio of temporally appropriate days covered by the prescription fills over the duration of the observation period.

There are additional adherence measures, such as Medication Refill Adherence (MRA) and MEDSUM, which are mathematically equivalent to methods used in calculating MPR. Another method is to use the Compliance Ratio, which is just yet another variation on the denominator calculation for MPR [4].

### 1.2.3 Gap-based adherence measures

Attempts have been made to bridge the divide between early and late stage adherence. One such measure is the New Prescription Medication Gap (NPMG). This measure is de-

defined as the proportion of days without sufficient supply from the date of the initial prescription to the end of the observation period (or censoring date if therapy is discontinued) [10]. This is operationally similar to either MPR or PDC depending on how oversupply is handled, although the number reported is  $1 - \text{MPR}$  (or  $1 - \text{PDC}$ ). The major difference in the calculation of NPMG is the definition of the start of the observation window, which will account for the time it takes for a patient to fill the first prescription. The end of the window is also fixed, rather than allowed to vary between studies like MPR or PDC. The NPMG captures patients who never started therapy (e.g. primary non-adherent). An additional strength of NPMG is that person-time can be censored if the prescriber switches or discontinues therapy and documents those orders in the EHR [4]. This measure also handles instances with few refills better than MPR or PDC methods. As implied by the name, NPMG is primarily useful for assessing adherence in treatment naive patients, rather than in those who are continuing therapy [10].

Other gap-based measures are Continuous Measure of Medication Gaps (CMG) and Continuous Multiple interval of OverSupply (CMOS) [4], both of which are similar to the converse of MPR. Additionally, there is the Continuous, Single Interval Measure of Medication Gaps (CSG), which looks at the number of days without any medication in the interval divided by the total time of the interval [11].

#### 1.2.4 Measures of persistence

Whereas adherence refers to how well the patient implements the prescribed regimen, persistence refers to how long patients stay on treatment [12]. One operational definition of medication persistence is the duration of time from the initiation of therapy until discontinuation. In order to analyze persistence, a limit on the number of days allowed between refills should be prespecified. This time frame is referred to as a permissible gap. The determination of what constitutes a permissible length varies with the drug and the treatment situation. A permissible gap should be defined as the maximum amount of time

that a patient could go without medication and not anticipate reduced or suboptimal outcomes. Persistence is reported either as a continuous variable in terms of the length of time (in days) for which the medication was available to the patient, or as a dichotomous outcome denoting whether the patient was persistent within the predetermined time frame [2]. Some researchers conflate persistence with adherence, and use metrics such as suboptimal secondary adherence as indicators of non-persistence (e.g. MPR < 80%) [13]. Finally, persistence is commonly operationalized as a time-to-event variable and analyzed via survival analysis [14].

The distinction between early and later stage persistence measures is not standardized across studies. While later stage persistence can be reported as the number of days on continuous therapy, early stage persistence is reported as a dichotomous variable indicating whether the patient is still persistent up to a prespecified point. Raebel et al. defines early stage persistence: “A new prescription was dispensed (Primary Adherence) and at least one refill of that prescription was dispensed over a time period consistent with (implying) current use of the drug” [4].

When persistence is reported as a dichotomous value (i.e., persistent versus non-persistent), often the criteria for the classification is not standardized. In a review of 58 studies of medication persistence, the allowable gap was highly variable, ranging from 7 to 180 days, the median being 30 days. Such alterations of what is deemed allowable can lead to drastic differences in persistence measures [13].

Alternative methods for measuring persistence include: counting the number of prescription refills over a defined time period [15], and taking the proportion of patients who filled a prescription within the last 60 days of the study period [16].

### 1.2.5 Trajectory measures

The metrics described above provide a cross-sectional summary of the data, which assumes that patient behavior is unchanging over time. Those methods fail to capture the

longitudinal aspects inherent in the data, when it has been shown that adherence to medication changes over time [17]. Rather than summarizing adherence over the whole study period, it can instead be calculated from one dispensing to the next [4], which confers the advantage of simultaneously capturing both adherence and persistence [15]. This calculation allows for longitudinal data analyses to be employed in characterizing adherence patterns. One such method that has recently been applied is to use group based trajectory models (GBTM). The GBTM is a statistical method that is designed to identify a finite number of groups of individuals following similar trajectories over age or time of a single outcome or behavior [18].

### 1.3 Efforts towards standardization

There have been efforts to summarize and standardize measures of medication adherence. A notable study investigated properties of Continuous Multiple interval measures of medication Availability (CMAs) [19]. This term refers to the large class of measures of medication adherence calculated by taking the ratio of drugs obtained within some observation period divided by the length of the observation period [20]. Vollmer et al. identified nine variants based on different study aims and assumptions. They termed these measures CMA1 through CMA9, and reported the medication adherence ratio calculated by each, using a large population ( $N = 6093$ ) of patients prescribed respiratory medication over a 15-month period. The nine measures mostly differ based on how the end of the observation window is determined (e.g., whether to use the study end date, or the date of the last fill), how the start of the observation window is determined (e.g., whether to use the date of the first fill, the study start date or the date medication was first prescribed), whether or not to cap the ratio at 100%, and whether oversupply should be accounted for in a time-forward manner. The list of different calculations is not exhaustive, but it does afford insight into the fact that many of the named measurements are all variations on the same theme, and that the inherent assumptions of each can significantly alter medication adherence outcomes.

Two of their measures, CMA7 and CMA8, do not require patients to have filled any prescriptions, and thus allows for the entire study population to be included in the analysis. These variants are provide a composite of both primary and secondary adherence (similar to NPMG).

In addition to looking at distributions of adherence ratios over each interval, they also looked at adherence over an elongating window: each of the first 3, 6, 9, 12 and 15 months of the study period. They report that shorter observation windows result in higher CMAs, which stabilizes at 9 months, suggesting a bias for patients with less time in the study. The use of the term “bias” may not be warranted here, however, as the number reported accurately reflects the patients’ adherence based on the prespecified conditions. Patient forgetfulness, attrition, and drop-outs accumulate with time and therefore, one would not expect adherence ratios at 3 or 6 months to mirror what is seen at the end of a longer observation period.

They do not recommend one CMA over the others, declaring that the choice of which adherence measure to use is based on the richness of the data, the chronicity of the disease under study, the availability of other therapies, knowledge about standards of practice, and finally, the scientific question addressed by the study. They note “No single measure is likely to be optimal for all occasions” [19].

### 1.3.1 AdhereR package

As part of this recent push to further standardize adherence research, Dima and Dediu [14] have created open-source software via the R programming language [21]. The package, entitled AdhereR, allows for a flexible and comprehensive investigation of EHR-based adherence to medications. The software includes many highly-parameterized functions, which allows researchers the flexibility to suit the needs of their studies, and is based off of the CMA framework as outlined by Vollmer, et al.

In addition to adherence, the AdhereR package can estimate persistence measures as



defined by a treatment episode duration. A treatment episode is a period of active medication use. Two consecutive medication events are considered to belong to the same episode if the time between the start of the second and end of supply from the first does not exceed a researcher-defined permissible gap length.

Interactive and publication-ready plotting functions allow for visualization of medication events. These plots allow for exploration of longitudinal medication use patterns, as well as providing a side-by-side comparison of the impact of different calculation methods.

The overall aim of the AdhereR software is to allow researchers to better understand the data, select clinically-meaningful study parameters, document the decision-making process, and communicate this entire process in a transparent and reproducible manner [14].

#### 1.4 Further considerations and limitations

Any method that utilizes claims data to measure adherence is a surrogate measure for actual medication taking, hence it is a crude measure. It is based upon the assumption that the act of filling a prescription is the same as taking all of the prescribed medication in the correct manner. To this end, many of these measures are estimating a best-case scenario for adherence to treatment. In determining the best method to calculate adherence and persistence, it would be best to be as conservative as possible within the framework of the assumptions and decisions being made.

Many of the above methods of characterizing adherence and persistence are subject to the same stipulations. Buono et al. suggest that these measures more reliably handle chronic, rather than acute, treatment regimens, and are less reliable for non-oral medication where a single dose is difficult to quantify [5]. However, with a proper accounting for end-of-regimen induced censoring, acute treatments can conceivably be measured as well as chronic regimens. Furthermore, not all non-oral medication is subject to quantification issues. In fact, one could postulate that a 30-day injectable is subject to the exact same (or perhaps less tenuous) assumptions as a 30-day supply of pills: once the medication is in the

patient's hands, whether it is being administered properly can not be measured with EHRs.

Time the patient spends in the hospital or as a resident in a long-term care facility, should be accounted for in a consistent manner. Some studies have addressed this by incorporating a grace period, excluding hospitalized patients, or by determining the number of days the patient was hospitalized and adjusting the measure of adherence accordingly [22].

The greatest decline in persistence for many chronic medications occurs within the first year of initiating a new therapy [17]. Studies evaluating a population of new users would be expected to find very different estimates of adherence or persistence than a similar study with a population of chronic users, even if the same definitions and methods to evaluate adherence and persistence are employed [22]. Therefore, it is of high importance to characterize the cohort in terms of new users and chronic users, and perhaps distinguish between the two when reporting results.

There are three types of drug use to consider when calculating adherence. The first is simple drug use, which one medication of interest per patient. The second is defined as drug switches, which occurs in patients who start on one therapeutic agent within the observation window, then switch to a different medication in the same class and never refill the first drug. The final type is multiple drug use within a therapeutic class (also known as polypharmacy), which is when the patient is prescribed multiple medications to be taken concurrently [8]. Calculating adherence becomes increasingly complex with drug switching and polypharmacy, and thus clear documentation of the methods used should be included when reporting the results.

Another major potential for bias and/or confounding in measuring and comparing adherence among patients is the variable number of days supply dispersed at each fill. Depending on factors such as the health care system, insurance company, or even the medication itself, drug supplies can vary widely from patient to patient, or even within the same patient. Consider a patient who receives a 90-day supply versus one with a 30-day supply. The second patient will need to complete the act of filling a prescription three times in order

to achieve the same fill volume over the 90-day period; this means a greater potential for being non-adherent. Taitel et al. report an MPR of 14% lower for patients with a 30-day supply versus those with a 90-day supply [23].

To compound an already imprecise measure of medication adherence, a common method for reporting adherence is to define a cut-point (usually 80%) for MPR or PDC, such that patients achieving a rate higher than this number are deemed adherent, and those below non-adherent. However, there are only a few medications for which a clinically investigated cut-point has been determined, such that patients above that threshold have little to no expected decline in health outcomes [15].

Finally, there is controversy about how to report measures of medication adherence. Some researchers will report primary non-adherence or early non-persistence with the study-level average, while others only report summaries for patients with secondary adherence of later-stage persistence. The latter method is argued as substantially inflating estimates, as it distorts the true relationship between medication adherence and clinical outcomes [24].

### 1.5 Unification of adherence measures

Of the four general measures discussed (primary and secondary adherence, and early and later stage persistence), the greatest need for standardization lies within secondary adherence. Both primary adherence and early stage persistence are clearly defined. Standardization of later stage persistence requires more consideration, much of which spins directly out of the discussion of secondary adherence. Secondary adherence and later stage persistence are complimentary measures in that together they provide a clearer picture of the degree to which patients are compliant with treatment.

Before proceeding, it is important to understand what adherence is, and what it is not. Adherence could better be thought of as “adherence potential,” because once the prescription is filled, the patient’s behavior is completely masked to researchers. Any actual med-

ication taking is assumed. Adherence measures are a summary statistic that obscures the chronology of patient behavior, and it also condenses the magnitude of adherent days into a ratio, therefore concealing the length of time the patient was observed.

Most of the discussed adherence measures are based on the same general calculation: the total number days supply of a medication divided by the number of days in the observation window. Other variants on that general theme is to measure days not covered by medication over various observation windows. We will consider only secondary adherence from here on, and define gap-based measures as measures of non-adherence. Unless the aim of the study is to measure medication acquisition rather than medication adherence, the PDC method of carrying oversupply forward is preferred to taking an overall total of days supply. A simple example illustrates this preference: suppose over a ten day observation period a patient obtains ten days worth of supply on the tenth day. By the MPR calculation, this patient will be considered to be 100% adherent, whereas with the PDC method, just 10%. As illustrated, the PDC method of only carrying oversupply forward is less susceptible to bias than MPR. In the following sections we attempt to unify the measurement of adherence within the PDC framework, while emphasizing the importance of transparency in study methodology. To reduce confusion of terminology, we introduce the term Medication Adherence Potential (MAP) in the following discussion of how to best measure adherence.

As with any adherence measure, unless the researcher has a clear understanding of how to define the study parameters, confusion is likely to follow. We define and make recommendations for those parameters, beginning with the denominator calculation.

### 1.5.1 Denominator calculation

The denominator consists of two components, the start and end of the observation period. The start of the observation period is typically the date of the patient's first fill within the study period. This convention is clear for treatment naive patients. For patients who are

continuing therapy, however, the researcher must decide whether to allow for a look back into prescription fills prior to the beginning of the study period. This would give a more informative summary of adherence, although it would not be advised in scenarios where a new intervention is being administered, as the patient would not have experienced the intervention prior to the onset of the study period. (Note that such a scenario could easily occur. For example, the intervention happens on January 1 while the patient's next refill date is January 15.) One method which could ameliorate the dilemma of defining a start date would be to use the day after the last covered date of the first fill, or the date of the second fill, whichever comes first. This solves two problems: i) patients are not immediately awarded 30 days (or whatever the fill volume is) worth of adherence to their numerator calculation; ii) patients who are on medication prior to the study period would be given a start date within the study period. It also has the additional benefit of creating a distinction between primary and secondary adherence.

For the end of the observation period, a common recommendation is to use the end of the study period, and any oversupply carried by the patient at this point should be excluded from the numerator calculation. This is an obvious choice for the end of the observation period for patients who are on study for the entire duration. An assumption of this definition is that no patients have experienced an event that would preclude their being on study or remaining on treatment, such as death or moving away. Therefore, using the end of the study period may bias the study-level MAP in a negative fashion. One method to mitigate this potential bias would be to define the end of the observation window as either the end of the study window for patients followed for the whole study, or the date the last fill is exhausted for patients who are not. This is likely to over-estimate the true MAP, as some of the patients will be non-adherent while others will be censored. If censoring information is known, this rule could be applied conditionally, and this method would provide the correct end of the observation window for every patient. However, censoring information may not readily be obtained. A third option would be to use the date the last fill is exhausted plus

a set amount of grace days as the end of the observation window for those patients who discontinue medication.

A way to circumvent the ambiguity of patients who discontinue medication would be to define the end of the observation period as the day before the date of the last fill within the study period. This method in conjunction with combined with using the day after the last covered date of the first fill has the benefit of restricting the observation period to the duration of time in which the patient has control over his or her adherence. Combining these two methods would require the patient to have at least three fills.

### 1.5.2 Numerator calculation

Once the observation window is set, calculating the number of days covered is carried out just as with PDC, which utilizes the following algorithm: i) generate a supply diary to represent every day within the observation window; ii) at each instance of a prescription fill (with N days worth of drug supplied), allot one day's supply to each of the next N dates within the supply vector; iii) if one prescription window overlaps with another, shift the latter prescription window until the two covered periods no longer overlap; and iv) sum up the number of covered days in the supply diary.

From there, MAP is simply the ratio of the numerator and denominator, i.e., the number of days covered over the length of the observation window. Any oversupply is to be carried forward in time only. Regardless of how much oversupply a patient carries, days supply is not to exceed the length of the observation window, preventing MAP from being greater than 100%. Refer to Figure 1.1 for an example using real data.

### 1.5.3 Drug switching

The above description is based upon the assumption of only one drug per patient, with no switching of medications while the patient is on study. There are many scenarios in which a patient may be switched from one therapy to another, such as a change of insur-

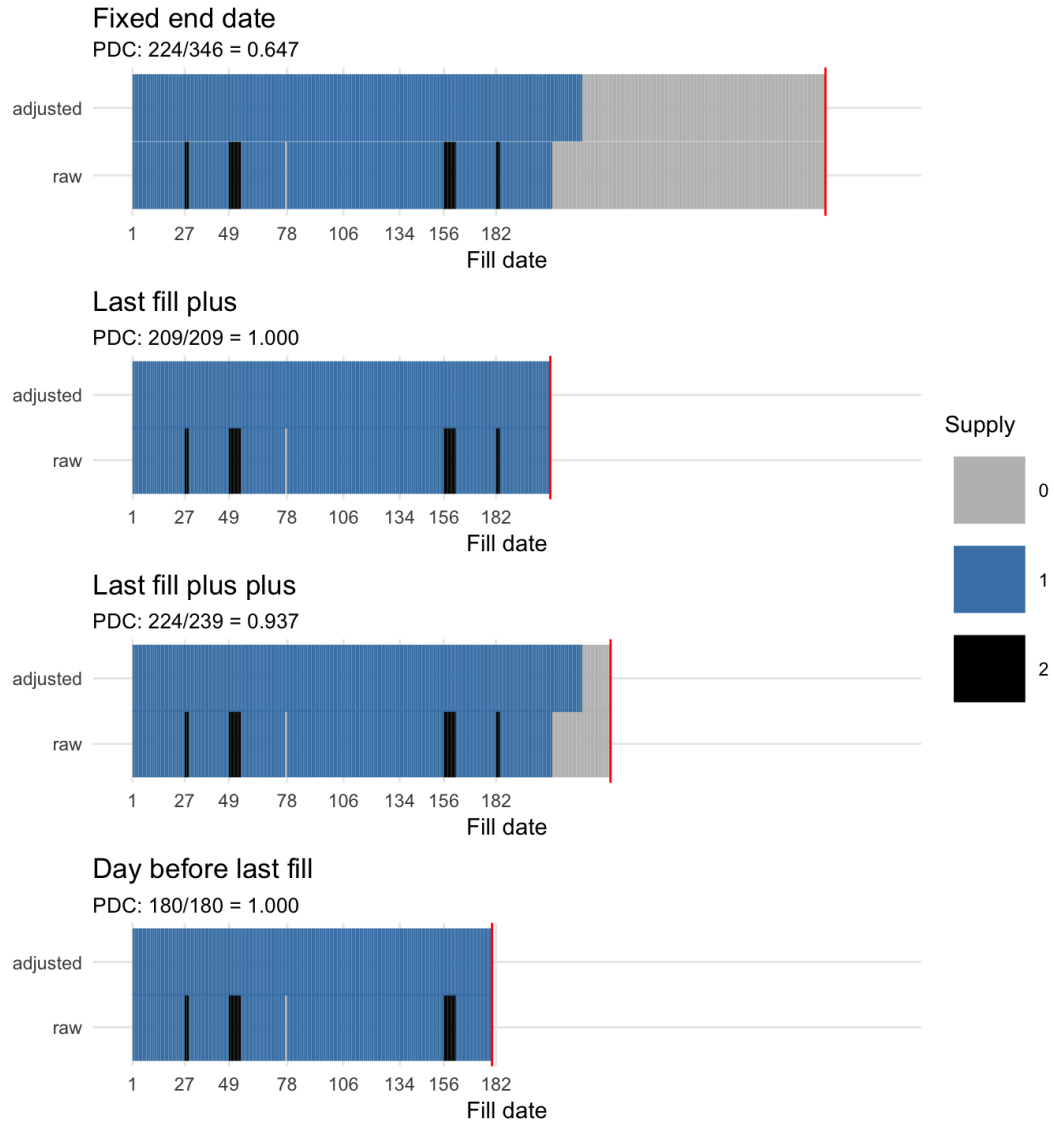


Figure 1.1: Calculation of PDC based on different end date rules. The **raw** barplots indicate the medication adherence supply diary, where black bars indicate a surplus of medication (caused by an early fill), while the grey bars indicate days where the patient is not covered by medication (gap days). The **adjusted** bar represents the supply diary after the oversupply has been shifted forward in time according to the rules for calculating PDC. Conceptually, the black bars can be thought to slide forward in time until a gap in medication occurs, then the excess supply provides coverage for otherwise uncovered days. Here we present supply diaries for the same patient according to four different end date rules. This patient had a gap in coverage at the end of the study period, thus the observation window is more variable than a patient who was followed through to the end. PDC calculations are displayed above each plot. Note that day before last fill and last fill plus methods arrive at the same PDC result, but with different lengths of the observation window.

ance or treatment being deemed ineffective. A common definition of a drug switch is the dispensing of a different drug within the same class at any point during the observation period.

Methods for how to measure medication adherence in instances of switching are not standardized, and should be chosen depending on the study objective. Because study objectives may vary, the researcher will need to define what constitutes an appropriate alternate medication. For instance, does a change in dosage or a change from brand name to generic constitute a different treatment event? If the two medications are comprised of the same chemical molecule, then could it be argued that the current treatment episode should be considered ongoing? However, if the patient is switched to a drug with a different formulation or mechanism of action, the decision to differentiate between treatment episodes would be a study-specific decision. If the research question at hand was patient adherence to types of medication, two treatment episodes should be considered. This will create an additional record in the study with a different adherence measurement for each medication regimen by every patient who switched therapy. Patients with multiple measurements require special attention, as adherence rates can no longer be considered independent within patients. On the other hand, if the adherence metric of interest is on the level of treatment, then the switch should be considered as ongoing adherence. Thus, when drug switching is present, the researcher will need to determine whether the treatment or the drug is the level of interest.

In either of the cases above, the question of how to handle oversupply should be addressed as well. There are two likely scenarios: i) the patient was instructed to stop taking the first medication either immediately, or upon filling the prescription for the new medication; or ii) the patient was instructed to finish the first drug supply before starting with the second. Assuming treatment level adherence is of interest, oversupply would be added into the numerator calculation, however, if the drug is the level of interest (multiple treatment episodes), then oversupply of the first drug should be removed from the numerator



calculation.

#### 1.5.4 Multiple concurrent medications

The many possible sources of confusion and bias discussed above are further compounded when the study takes multiple concurrent prescriptions into account. Some methods for measuring adherence for polypharmacy include: averaging the adherence to each individual drug; using the number of days with at least one drug in the regimen available in the numerator calculation; and a daily polypharmacy possession ratio (DPPR) [25]. The method for DPPR is as follows: “Look at each day in the observation period separately, and determine how many medications are available, set a score between 0 (no medication available) and 1 (all medications available) weighted by the number of medications to be taken each day, resulting in daily scores indicating the proportion of medications available for each day. Sum the scores and divide by the number of days in the observation period to obtain the proportion of all medications available for daily use [25].” DPPR can be thought of as a daily weighted average for the number of medications the patient is on.

All three of the above methods are problematic for the same reason, as evidenced by the following example. A patient is prescribed a medication regimen of five different drugs to be taken concurrently over the course of a year. The patient has complete adherence to four of the five drugs, but never fills the prescription for the fifth. Under each of the three listed methods, this patient would have a MAP of 80%, 100% and 80%, respectively, and thus would be deemed adherent at the 80% threshold. By no means should this patient be considered adherent to the total prescribed treatment.

Any composite measurement is susceptible to the scenario described above. A more straightforward way of handling polypharmacy would be to simply report individual MAPs for each drug. This affords the researcher a much clearer picture of the actual adherence patterns of the patients in the study. Upon looking at the individual MAP values, if there is no significant difference among the adherence potential for the different drugs, the re-

searcher can then take a weighted average of the multiple measures.

### 1.5.5 Additional considerations

Discontinuation of a subset of the regimen, switching, or starting a new medication while already on study can theoretically be handled by the described methods as long as the researcher has clearly defined the parameters of the study. Gaps in medication possession due to hospitalization could also be accounted for by excluding any time that the patient was unable to access his or her own medication supply due to time spent in the hospital. In practice, this may be as simple as subtracting the number of days spent in the hospital from the length of the observation window, although this is subject to many assumptions that should be considered depending on goal of the study.

The MAP (which is just a multi-purpose method for calculating PDC) is not a catch-all measure of patient adherence. While it is superior to measures such as MPR due to its ability to incorporate the timing of the fills, the nature of collapsing it into a simple ratio masks a lot of information. For instance, a patient with 90 days supply in a period of 120 days has the same MAP as someone with 300 days supply in a 400 day period. The latter would provide more information regarding the patterns of behavior exhibited by the second patient, but MAP alone cannot convey that information. In addition, it does not address the fact the different amounts of medication can be prescribed to different patients, or to the same patient at different times.

### 1.5.6 Reporting MAP

When reporting MAP, it is of critical importance to clearly define the parameters that went into the calculation. Without transparency, making comparisons across studies would be futile.

Depending on the start date rule applied, the MAP can encompass both primary and secondary adherence. However, in rebuttal to Raebel et al. [24], we assert that it is more

desirable to keep primary and secondary measures distinct, as they are really describing two different populations. Rather than providing a composite measure of primary and secondary adherence, it would be better to present summaries for both measures (e.g., “70% of the cohort filled the first prescription, and among them the average MAP was 85%”).

When the study-level MAP is reported, it is often calculated as a raw mean of each individual MAP. This method introduces bias and gives undue weight to patients who were on medication over a short period of time versus those with a much longer chronology of adherence. Rather than a raw mean, study-level MAP should be reported as the sum of all patients’ days worth of supply over the sum of the length of all patients’ observation windows. This method accounts for person-time on the medication and would not be unduly influenced by outlying patient MAP rates.

Other metrics to report along with MAP should include the numerator and denominator, the number of fills, the average days supply per dispensation, the number of gaps, the total length of the gaps, the number of non-permissible gaps and the total length of non-permissible gaps.

## 1.6 Classification of persistence measures

Measures of persistence are more disparate than those of adherence, and are therefore more difficult to unify under one general framework. The various measures that have been used to describe persistence can be summarized in two distinct categories: time until discontinuation and a yes/no indicator of persistence based on specific criteria. The main issue with either method is that the definition of non-persistence is highly variable based on what criteria is assumed [26]. Furthermore, many studies conflate adherence and persistence, often using a value of MPR or PDC above a certain cut-point as the definition of persistence.

Persistence measured as time until discontinuation is subject to the same problem as

MAP in terms of defining the end of the observation window. Unless information about stop orders, death, and transition to new pharmacy systems is available (i.e., patient censoring information), analyses using survival methods may be biased. However, if censoring information is available, then a time-to-event analysis can suitably assess patient discontinuation patterns, subject to the definition of a permissible gap. The problem with allowing for a permissible gap is that it is often not clear what happens in the event of a non-permissible gap. In a scenario where a ten day lapse is deemed permissible, how would persistency be defined for a patient who has six months of continuous coverage, then an eleven day lapse, followed by six more months of continuous coverage? Persistent for six months only or two persistence episodes would both be viable under the definition of persistence. The AdhereR package can effectively compute multiple treatment episodes; however, the statistical method to model this measure is unclear.

Much like MAP, persistence can be confounded by the number of fills a patient has, or the quantity of pills dispensed. We present an alternative to persistence: Medication Refill Vigilance (MRV). The MRV is a summary of the number of times that the patient refills the prescription before their current supply runs out, divided by the number of refills the patient had during the observation window. This can be further generalized by altering the definition of what constitutes an allowable amount of time to pass between one prescription being exhausted and the start of the next. Instead of looking at the length of the medication lapse in days, one can instead set a minimum threshold based on an interval-based MAP calculation. The intent behind the MRV metric is to constrain the analysis to behaviors that can be directly observed (filling a prescription) while omitting those that can only be assumed (any actual medication taking). Rather than using the MRV ratio in statistical analyses, each instance of prescription filling can be used in a longitudinal analysis.

One advantage MRV has over MAP-based measures is that patients with large gaps are not penalized or misrepresented. One large gap in an otherwise impeccable record of medication filling behavior could skew an otherwise adherent patient's MAP to be low.

This is especially beneficial in closed health systems where out of network prescription fills do not appear on EHRs, or to account for periods of relapse or remission.

## Chapter 2

### Statistical Methods

#### 2.1 Introduction

In addition to a lack of consensus in measuring and reporting adherence, there has not been much agreement how adherence data should be analyzed. The analysis methods reported in the literature include ordinary least squares (OLS) [27], generalized linear models (GLMs) with a logit or a gamma link and/or a hurdle component [28]. Our aim is to find the best method for analyzing adherence data. We consider three methods: logistic regression, ordinal regression, and negative binomial regression. Additionally, we investigate the applicability of longitudinal data analysis using generalized estimating equations (GEE).

While patient adherence expressed as a ratio summarizes adherence, statistical analyses using the ratio could be problematic. The major reason can be illustrated in the following example: a patient with 30 days covered out of a 40 day observation period has the same MAP as a patient with 300 days covered out of 400. Although these patients have drastically different adherence profiles, they will contribute the same amount of information if the ratio is used in the analysis. Secondly, the use of ratios in regression models can lead to incorrect and misleading inferences [29]. The two driving factors for using a ratio as an outcome variable are either on the grounds of simplicity or itself being the quantity of interest based on the rationale that a ratio adjusts for the effect of the length of the observation window (i.e., the denominator value). However, as illustrated in the example above, this does not achieve the intended result. There are alternative methods available to adjust for the variable length of the observation window, some of which will be discussed later.

## 2.2 Logistic regression

Logistic regression is a typical method employed for modeling patient adherence to medication [30][31][32][33]. This is accomplished by dichotomizing a continuous measure of adherence, and classifying patients above some threshold (usually 80%) as adherent and those below as non-adherent. There are numerous reasons why dichotomization should be avoided [34]. Most importantly, much of the information will be lost. Such a method of classification implies that patients just above the 80% threshold are expected to have different clinical outcomes compared to a patient just below the threshold. This also assumes that patients far below the threshold exhibit the same patterns of behavior as those just below it. While we do not advise modeling adherence by dichotomizing the outcome, it will be included for comparison purposes.

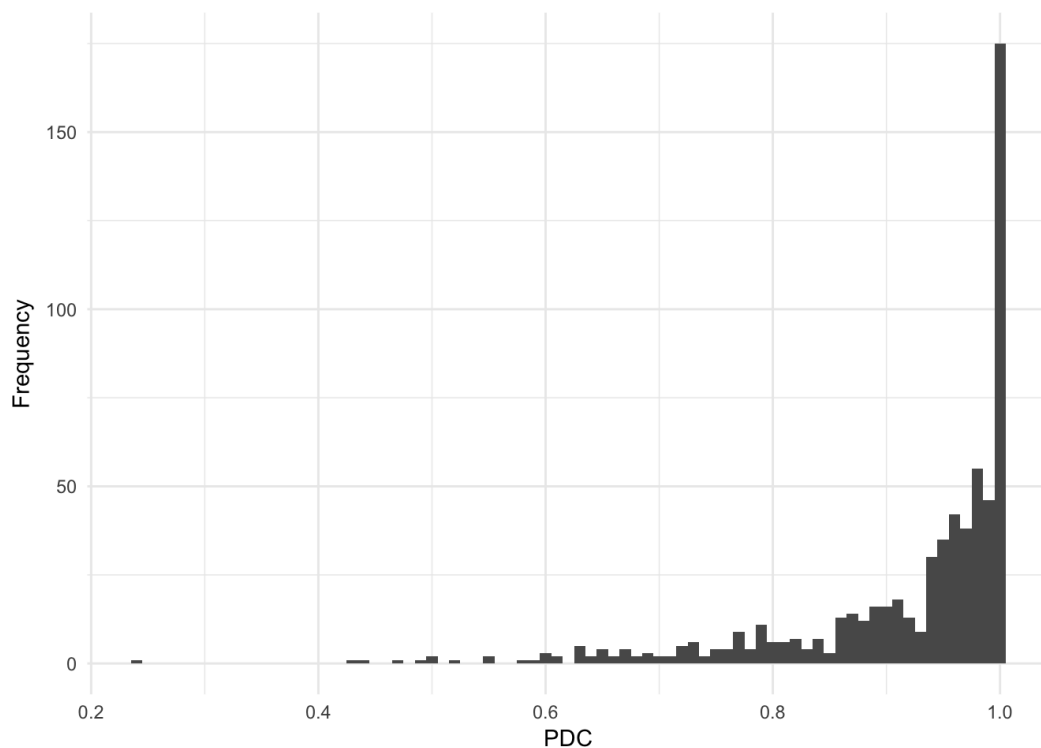


Figure 2.1: Histogram of PDC, calculated using the last fill plus method.

Selection of a cut-point should be clinically relevant, and there are very few medications for which this has been determined [15]. Performing logistic regression with an assigned

cut-point likely persists for the following reasons: historically being used, a lack of serious statistical consideration, or challenges present in the data.

In the dataset used as the case study in this thesis, and similar datasets we have analyzed, a high percentage of patients were 100% adherent, resulting in a skewed distribution of PDC (Figure 2.1). Such a distribution violates the assumptions of ordinary least squares regression, and any transformations on the outcome would not address the high density of patients with complete adherence (i.e., 100% PDC). In the following sections, we present alternative models to address this uncommon distribution.

### 2.3 Ordinal logistic regression

The cumulative probability ordinal model is a robust semi-parametric regression approach with several advantages over OLS. Unlike the linear model which assumes a normal distribution for  $Y|X$ , the conditional distribution of  $Y$  on  $X$ . The conditional distribution in ordinal regression need not be normal, or even continuous [34]. This is particularly advantageous when the distribution of adherence measures would be highly skewed.

Ordinal models are based on the ranks of the  $Y$  values [34]. Furthermore, this method is robust to outliers (e.g., patients with very low adherence relative to typical values).

Cumulative probability models can be constructed with various link functions; here we consider the logit link. An ordinal logistic regression, or proportional odds (PO) model, can be described as follows: let the ordered, unique values of  $Y$  be denoted as  $y_k, k = 1, \dots, K$ , and the intercepts associated with each  $y_k$  be  $\alpha_1, \dots, \alpha_K$ , where  $\alpha_1 = \infty$  because  $P[Y \geq y_1|X] = 1$  [34]. The general formula is given by

$$P(Y \geq y_k|X) = \frac{1}{1 + \exp[-(\alpha_k + X\beta)]}.$$

This formulation of  $Y \geq y_k$  makes the model coefficients consistent with the binary logistic model; that is, when  $Y$  can only take on two values, the interpretation of the ordinal



model is the same as a logistic model. For fixed  $k$ , the model is an ordinary logistic model for the event  $Y \geq y_k$ . The coefficients of  $X$  are log odds ratios, and a common log odds ratio is assumed for all events  $Y \geq y_k$  [34].

There are several assumptions in fitting a PO model, and the primary one is implicit in the name: the odds of the response being above any one cutoff point are proportional for all cutoff points. For each specific cutoff, the model has the same assumptions as the binary logistic model. The log odds of being  $Y \geq y_k$  is linearly related to each  $X$  and there is no interaction between the  $X$ s. Also the regression coefficients are independent of the cutoff level for  $Y$ , i.e., there is no  $X \times Y$  interaction if the proportional odds assumption holds [34].

Like how some of the basic assumptions of OLS are often violated, the proportional odds assumption can be violated as well. However, the PO model still can be a powerful model in this situation.

Ordinal regression has additional desirable properties in addition to those described previously. The model allows for estimation of the mean, estimation of quantiles as efficient as quantile regression (given that the PO assumption holds) and direct estimation of  $P(Y \geq y|X)$ . The latter property allows for the calculation of exceedance probabilities, which can achieve what is desired by dichotomizing the data, but without the egregious waste of information and power.

Ordinal regression makes a good candidate model for adherence data based on the highly skewed distribution of the data, and its ability to provide exceedance probabilities. However, it has some limitations when applied to this data that can not be overlooked. Because the outcome is expressed as a ratio, the proportional odds assumption is unlikely to be met. In addition, an overall ratio conceals the length of the observation window. Patients with 60 days covered out of 90 days will be modeled the same as patients with 120 days covered out of 180.

## 2.4 Negative binomial regression

Instead of using the proportion of days covered as the outcome, we will also consider using the number of gap days as the outcome. A ratio of 1 is equivalent to the difference between the denominator and numerator being 0, that is, the number of gap days will be 0. We will use the difference between the denominator and numerator calculation as an alternate outcome variable, which can be represented as gap days and hence considered as count data.

Based on datasets we have examined, we observed a high proportion of patients with zero gap days over the duration of their observation window. Also, depending on the rules for determining the observation window, there is potential for a very long-tailed distribution. Thus, it is unlikely that the assumption of the mean-variance relationship for a Poisson model will be met. Therefore, we examine the negative binomial model that can handle overdispersed count data as an alternative to a Poisson model to fit non-adherence data.

Negative binomial (NB) regression is a GLM with a log link function where a response is assumed to have a NB distribution conditional on the predictors. The NB model is an alternative to a Poisson model, where the variance and the mean are not equal. The variance of the NB distribution is still constrained to be a function of its mean, but has an additional parameter,  $\theta$ , called the dispersion parameter which provides additional flexibility [35]. For a random variable  $Y$  from a NB distribution, the variance is given by:

$$\text{Var}(Y) = \mu + \frac{\mu^2}{\theta}.$$

Thus the variance is quadratically related to the mean, and as the dispersion parameter  $\theta$  grows large, NB converges to a Poisson distribution [35].

There are multiple formulations of the NB regression model, among which the most common one is based on the Poisson-gamma mixture distribution. This formulation allows

modeling of Poisson heterogeneity using a gamma distribution [36]. The Poisson-gamma mixture distribution is given by:

$$P(Y = y_i | \mu_i + \theta) = \frac{\Gamma(y_i + \theta)}{\Gamma(y_i + 1)\Gamma(\theta)} \left( \frac{\theta}{\theta + \mu_i} \right)^\theta \left( \frac{\mu_i}{\theta + \mu_i} \right)^{y_i},$$

where

$$\mu_i = t_i \mu.$$

The  $\mu$  parameter is the mean incidence rate of  $Y$  per the unit of exposure,  $t_i$ , and is interpreted as the risk of a new occurrence of an event over the course of the exposure period,  $t_i$ . In adherence data, exposure is the length of the observation window in days. From the results of this regression, we can estimate the average MAP ratios, based on the average number of gaps days and the average length of the observation window.

In addition to the assumption for the variance, another assumption is that the coefficients are additive on the  $\log(Y)$  scale and that incidence rate ratios (IRRs) have a multiplicative effect in the  $Y$  scale [37].

One drawback to the NB regression model is that it is not recommended for application on small sample sizes. Other models that can handle overdispersion are available, which we do not think would provide as suitable a model as the NB model. Zero-inflated models have the ability to handle a large number of excess zeroes, which is founded on the assumption of structural zeroes. Implicit in that assumption is that some patients never have the potential for non-adherence. A quasi-Poisson method is also available to model overdispersion, but we preferred a NB model as it can take advantage of the full likelihood method, so that it can be used to directly compare with a Poisson model [37]. A NB is not an exponential family distribution, hence there is no canonical link, and a log link is customary to make it more similar to Poisson.

Count models have the additional benefit of being about to account for an offsetting

variable. The offset is an a priori known component to be included in the linear predictor. It is provided on the log scale, and represents the number of times that the event of interest could have occurred. Thus, count models are applicable to rates, or ratios, such as the number of gap days per days of observation.

Using the number of gap days as the outcome variable allows us to fit a data that matches the distribution of non-adherence data. One advantage over other models is that we are no longer constrained to using a ratio, and it can be argued that 20 gap days is objectively worse than 10 gap days. The variability in the length of the observation window (i.e., denominator days) can be accounted for by including an offset in the model. The disadvantage to this model is similar to that of other models described above, in that the ordering of the occurrence of gap days is not taken into account. For example, a patient with a single lapse of 20 gap days would be considered to have an equivalent outcome as a patient with 10 instances of two day gaps. The non-adherence behavior of these two patients could be considered to be very different, yet the model would be unable to differentiate the two.

## 2.5 Longitudinal regression modeling

Often adherence is a series of behaviors, and hence rather than modeling a summary statistic, it may be more informative to consider the recurrence of filling a prescription. To serve this end, we propose using GEE to analyze medication adherence data. The GEE can be used to estimate the parameters of a GLM with an unknown correlation between outcomes. The GEE is specified by a mean model and a correlation model, which means that there is a regression model for the average outcome, as well as a model for the longitudinal correlation. The marginal mean is given by:

for  $i = 1, \dots, N, j = 1, \dots, m_i,$

$$E[Y_{ij}|X_{ij}] = \mu_{ij}(\beta)$$

$$g[\mu_{ij}(\boldsymbol{\beta})] = x_{ij}\boldsymbol{\beta},$$

where  $N$  is the number of subjects,  $m_i$  is the number of fills for each subject,  $X\boldsymbol{\beta}$  is the linear predictor, and  $g(\cdot)$  is the link function as used in GLM [38].

One benefit to GEE is that, if the mean model is correctly specified, even though a working correlation model is incorrectly specified, under reasonable general conditions, consistency will be retained, although efficiency may be lost [38].

Within the GEE framework, the correlation structure is treated as a nuisance feature of the data, and only requires a selection of a working correlation model. Even with an incorrectly specified correlation structure, robust standard errors (SEs) will still be valid; that is confidence intervals will have the stated coverage. Correct specification of the correlation model can result in efficiency gains [38].

For adherence data, we will consider both the identity link and the logit link functions. We will use an exchangeable correlation structure, as it is reasonable to assume that within each patient, the medication filling behavior will be approximately constant over time.

Each repeated measurement for adherence data must be preceded and followed by a filling event, and therefore, information about the last fill is not used. This has the effect of both reducing the amount of fills for every patient, as well as eliminating the need to decide on a rule on how best to define the end of the observation window, as it by default must be the day of the last fill.

### 2.5.1 GEE for non-permissible gaps

Repeated binary logistic regression may be best suited for situations in which a non-permissible gap is clearly defined. At its simplest calculation, a non-permissible gap could be considered to be any instance in which the patient refilled their medication after the previous supply has been exhausted. Depending on the aim of the study, the definition of a permissible gap can be extended to any number of days. Furthermore, a non-permissible

gap need not be determined by the length of time without medication coverage, and can instead use a common threshold, such as an interval-based PDC (PDCi) below 80%. This has the effect of normalizing the proportion of time that patients are uncovered in the event that variable volumes of medication are dispensed.

Non-permissible gap data can be modeled using GEE with a binomial family link, using the binary outcome of filling on time (i.e., no gap or a permissible gap) versus not (i.e., having a non-permissible gap). This method affords us the ability to determine how well the patient complies with a recommended refill regimen, but just as with any other dichotomization of data, much information is lost. Furthermore, it has been previously shown that variable gap lengths can lead to varied inference [26]. Thus, this method is susceptible to many of the biases of the previously discussed models. Despite its shortcomings, this method still has the potential to provide insight into patient behavior.

### 2.5.2 GEE for continuous gap and surplus time

All of the outcomes of the above models are based on the notion that time without medication is the primary unit of interest. While there is no expected clinical difference in outcomes between people who refill on time compared to those who refill early, we might hypothesize that the set of behaviors or circumstances that lead to an early fill are different than those that lead to filling on time, or filling late.

Both PDC and gap days represent a truncation of the data. Both distributions are capped at what constitutes complete adherence. By looking at interval-based data, we find the capping mechanism to be somewhat arbitrary. Instead of looking at this truncated data, if we were to look at a composite of both gap and surplus days, the distribution is more normal and symmetric, which would allow for the use of a standard link function without performing data transformation. Thus, we propose a GEE model with an identity link.

Medication adherence is directly attributable to the behaviors, characteristics and circumstances that lead patient to refill their medication. When the unit of measurement is

restricted to a single refill interval, defined as the time from one refill to the next, a clearer picture of the data can be obtained. Interval-based calculations provide additional information including: the time between fills, amount of drug supplied at each refill, the timing of the refill (be it early, on time or late relative to the ostensible date of medication exhaustion), as well as the amount of oversupply carried over from the last refill interval.

The benefit of this method is that regression coefficients can be easily interpreted, as they are the mean difference in the average number of days from one fill to the next between the levels of the covariate of interest. One disadvantage of this method is that very large gaps can be common in the data, which maybe not be handled adequately.

## Chapter 3

### Case study

We begin this chapter with a description of real world data and the various methods of calculating the outcomes. Next, we will conduct a comprehensive sensitivity analysis using the methods outlined in the previous chapter, and discuss model checking.

#### 3.1 Data

The data consist of pharmacy records for 653 patients who received dispensations of medication from the Vanderbilt Specialty Pharmacy (VSP) to treat multiple sclerosis (MS). Prescription records for this study were restricted to the calendar year 2016. Patients were excluded from the study if they had fewer than 3 prescription fills with the VSP during the study period. A primary goal of for this study was to determine if patients who were new to therapy (i.e., treatment naive) had different adherence as measured by PDC compared to those who were continuing therapy.

##### 3.1.1 Patient demographics

The median age for study participants was 47 (interquartile range 40 - 56) and 75% of the cohort was female. The majority of patients (84%) were white, while 12% were black and 4% were either Asian/Pacific Islander, multi-race or did not identify. Of the 653 patients, 135 were starting treatment for the first time. Forty-two percent of the cohort utilized government-sponsored insurance, and 73% received one form of copay assistance. Eight different drugs were prescribed to patients (Betaseron, Extavia Avonex, Rebif, Plegridy, Copaxone, Aubagio, Gilenya, Tecifdera), with Copaxone and Tecifdera making up the largest portion of prescriptions with 23% and 21% of the cohort, respectively. Because of staggered entry times and variable prescription volumes, patients obtained a variable



Table 3.1: Patient demographics from the MS study

	<i>N</i> = 653
Age	40 47 56
Race	
Non-white	16.2% (106)
White	83.8% (547)
Sex	
Female	75.2% (491)
Male	24.8% (162)
Treatment naive	
No	79.3% (518)
Yes	20.7% (135)
Use of Copay Assistance	
No	27.3% (178)
Yes	72.7% (475)
Medication	
Betaseron	4.1% (27)
Extavia	0.9% (6)
Avonex	9.8% (64)
Rebif	9.8% (64)
Plegridy	3.7% (24)
Copaxone	22.8% (149)
Aubagio	10.4% (68)
Gilenya	17.5% (114)
Tecifdera	20.9% (137)
Fills	
3	6.1% (40)
4	8.6% (56)
5	6.3% (41)
6	5.5% (36)
7	5.5% (36)
8	5.8% (38)
9	5.8% (38)
10	6.9% (45)
11	10.7% (70)
12	17.8% (116)
13	14.7% (96)
14	4.8% (31)
15	1.5% (10)

*a b c* represent the lower quartile *a*, the median *b*, and the upper quartile *c* for continuous variables. Numbers in parentheses after percentages are frequencies.

number of fills. The most common number of fills was 12, which comprised 18% of the cohort. Complete demographic information can be found in Table 3.1.

Across the 653 patients, there are 6107 prescription records. The most common days of supply for drugs is 28 days (57% of prescriptions) and 30 days (38% of prescriptions) (Table 3.2). Patients were only prescribed one type of medication within the study period, thus instances of drug-switching or polypharmacy are not attributes of this data.

Table 3.2: Days supply administered at each fill.

<i>N</i> = 6107	
Days supply	
4	0.10% ( 6)
12	0.03% ( 2)
14	0.02% ( 1)
15	0.03% ( 2)
28	57.49% (3511)
30	37.89% (2314)
60	0.05% ( 3)
84	2.85% (174)
90	1.54% ( 94)

### 3.1.2 PDC calculations by different methods

As described in Chapter 1.5, determining the best method of calculating PDC is contingent on the parameters and the objective of the study. Because patients were enrolled continuously throughout the year of the study rather than all at the same time, it is better to use the day of the patient’s first fill as the first date in the observation window.

In determining the end of the observation window, there are four options that we will consider. The first is to use the end of the study period, December 31, 2016, for all patients, called “fixed end date”, or “Fixed”. This is the method recommended by the PQA guidelines, however, this assumes that all patients were expected to be on medication for the whole year. This was not the case with the MS data, as patients dropped out or changed pharmacy throughout the year. In a previous analysis of this data, the rule of selecting the

earlier occurrence of either the last fill being exhausted or the study end date was used, called “last fill plus”, or “LFP”. As discussed previously, this may bias estimates of PDC high, as it assumes that all patients lost to follow up experienced a discontinuation event rather than exhibited non-adherent behavior at their last fill. An extension to the LFP method is to use the earlier occurrence of the study end date or the date the last fill was exhausted plus 30 days, called, “last fill plus plus”, or “LFPP”. This allows for the possibility that people were non-adherent at their last fill, but caps this medication gap at 30 days, which is another artificial constraint that may not reflect actual medication adherence behavior. Finally, we can disregard the decision of a censoring rule altogether, and use the day before the last fill as the final date in the study, called “day before last fill”, or “DBLFF”. This could be a valuable method in that filling a prescription does not provide insight on future adherence, yet it does give credence to the assumption that the previous fill has been completed, or is near completion. Thus, each fill gives more information on the previous fill than the current one. As this method removes the last fill, the observation window is shorter.

In addition for determining the best method for defining the observation window, we will also investigate how the results would be impacted by defining the cohort in terms of complete follow up and the number of fills. Limiting the study population to patients who were covered at least through the last day of the observation window removes any ambiguity on how to handle patients who ostensibly have dropped out. Additionally, because three months may not be enough time to effectively assess a patient’s medication refilling behavior, as discussed in Vollmer, et al. [19], we will restrict the cohort to patients with at least six fills as well as at least nine.

Table 3.3 provides summary statistics of PDC based on the end date of the observation window with the entire cohort, or the definition of the cohort with the method of LFP. Of the four end date rules applied, the Fixed method results in the lowest mean PDC, which is expected as all patients who have been lost to follow up are counted until the end of the

Table 3.3: Summary statistics of PDC by the end date of the observation window with the entire cohort and the definition of the cohort with PDC calculated using LFP

PDC	By end date				By cohort		
	LFP (N = 653)	Fixed (N = 653)	DBLF (N = 653)	LFPP (N = 653)	Complete (N = 516)	Six fills (N = 516)	Nine fills (N = 406)
Mean	0.92	0.87	0.91	0.90	0.92	0.92	0.94
SD	0.11	0.18	0.12	0.12	0.10	0.10	0.07
Median	0.96	0.95	0.96	0.95	0.96	0.96	0.96
$N < 80\%$ (%)	90 (14%)	148 (23%)	99 (15%)	117 (18%)	61 (12%)	67 (13%)	29 (7%)

**LFP**: last fill plus; **Fixed**: fixed end date; **DBLF**: day before last fill; **LFPP**: last fill plus plus, **Complete**: patients who were followed through to the end of the study period; **Six fills**: patients with at least six fills; **Nine fills**: patients with at least nine fills

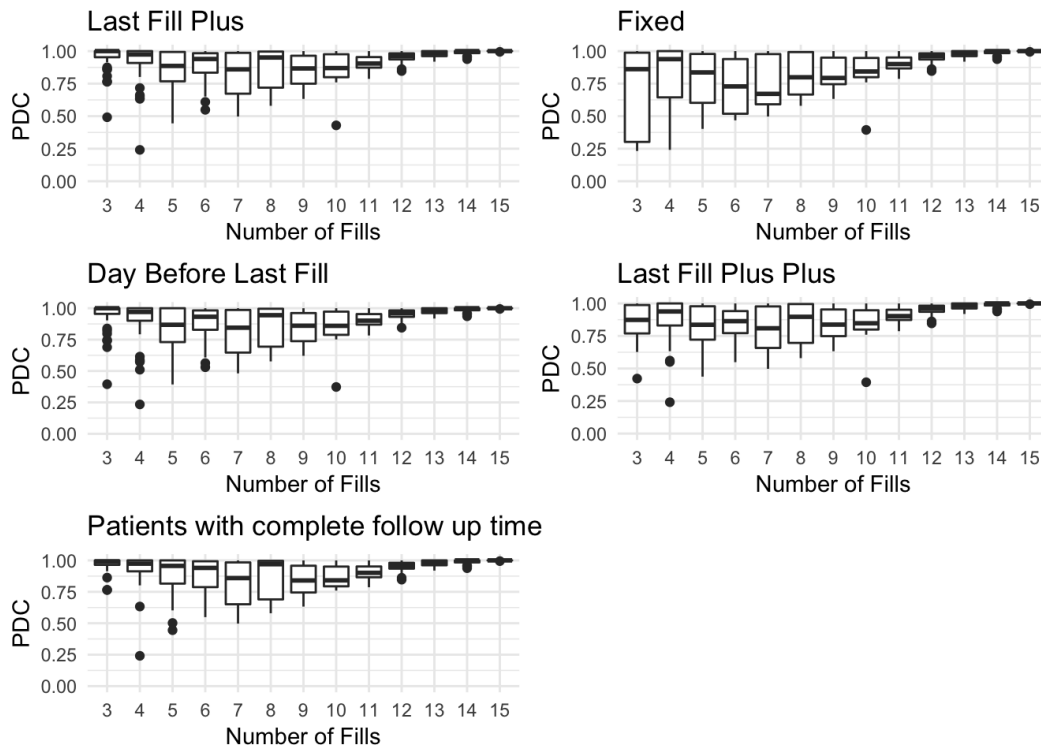


Figure 3.1: PDC comparison based on different end date rules. Patients with twelve or more fills have nearly complete adherence regardless of the rule applied. The greatest variability within each method is in the mid-range of the fill numbers, and the greatest variability among the different rules is for patients with three fills. Patients who could be considered to drop out after three fills could potentially have a long period of uncovered time (Fixed), or their uncovered time could be disregarded entirely (LFP). Patients with complete follow up have the least variable PDC values, as dropouts no longer need to be accounted for.

study period. The other three end dates give roughly similar results for mean and median PDC. In restricting the cohort to those who have complete follow up, the sample size is reduced to 516 patients. Restricting the minimum number of allowable fills to six achieves the same sample size, although they are not exactly the same cohort, and further restricting to nine fills reduces the cohort further. The cohort reduction without clear justification would not be recommended in practice, but this may provide insight into the properties of PDC in actual data. The LFP calculation was the default method applied to calculate the adherence statistics for the reduced cohorts.

### 3.1.3 Results of PDC calculations

The distribution of PDC calculated using the LFP method is displayed in Figure 2.1. There is a high proportion of patients with a PDC value of 1.0. This is not uncommon with PDC, which is part of the reason why applying statistical models to the data presents a challenge. In order to model PDC, it is better to fully understand the structure underlying this distribution.

Due to staggered enrollment and varying amounts of drug supplied at each refill, the number of fills that patients had throughout the study ranges from 3 to 15. Figure 3.1 displays boxplots of the PDC value for each of the number of fills based on the four end date rules applied, as well as the cohort who were enrolled through the end of the study period. Two things are notable in this plot: patients with a high number of fills have high values of PDC, and patients with a low number of fills also have a high PDC. The former makes sense in the context of the study, in that the duration of the study period is only 12 months, and most refills are for either 28 or 30 days. Thus a patient with a high number of fills will have a very high proportion of days within the study covered by medication. The latter is not as obvious. Upon a closer inspection of enrollment dates however, we can see in Figure 3.2 that there is a large uptick of patients with three or four fills beginning roughly in October. Those patients have a high PDC for the same reason as patients with

12 or more fills who started in January.

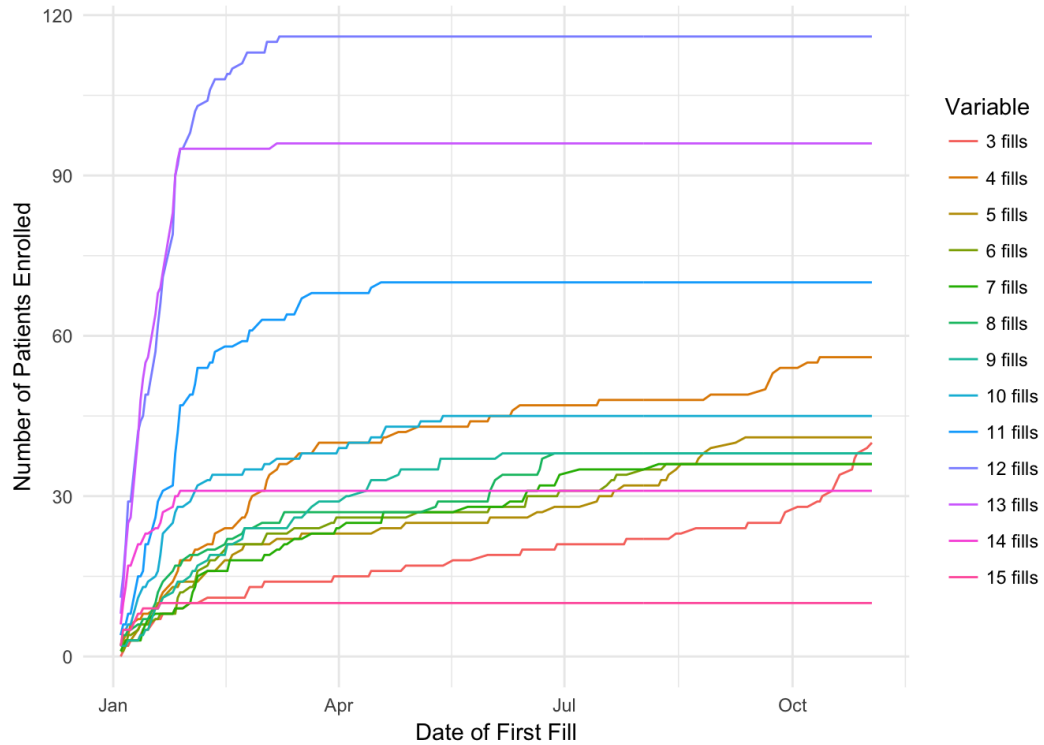


Figure 3.2: Cumulative number of enrolled patients by total number of fills. Patients with twelve or more fills must have an enrollment date early in the year to accommodate many refills, and as the number of fills decreases, the time at which the total number enrolled levels off. Patients with a small number of fills enrolled throughout the study period. Note that there is a sharp uptick of patients with three and four fills just before the end of enrollment.

### 3.1.4 Gap-based outcomes

The distribution of the total number of gap days is shown in Figure 3.3. Because gap days are calculated as the numerator minus the denominator in the PDC calculation, we expect to see the high density of patients with zero gap days. The tail of this distribution reaches beyond 250 days, which suggests that the patient in question either experienced a cumulative 252-day lapse in medication, or the patient filled the medication at a different pharmacy for the time in question.

In the MS data, there are 116 instances of patients having a gap between two fills of greater than one month long. The longest individual gap period is 252 days, with other

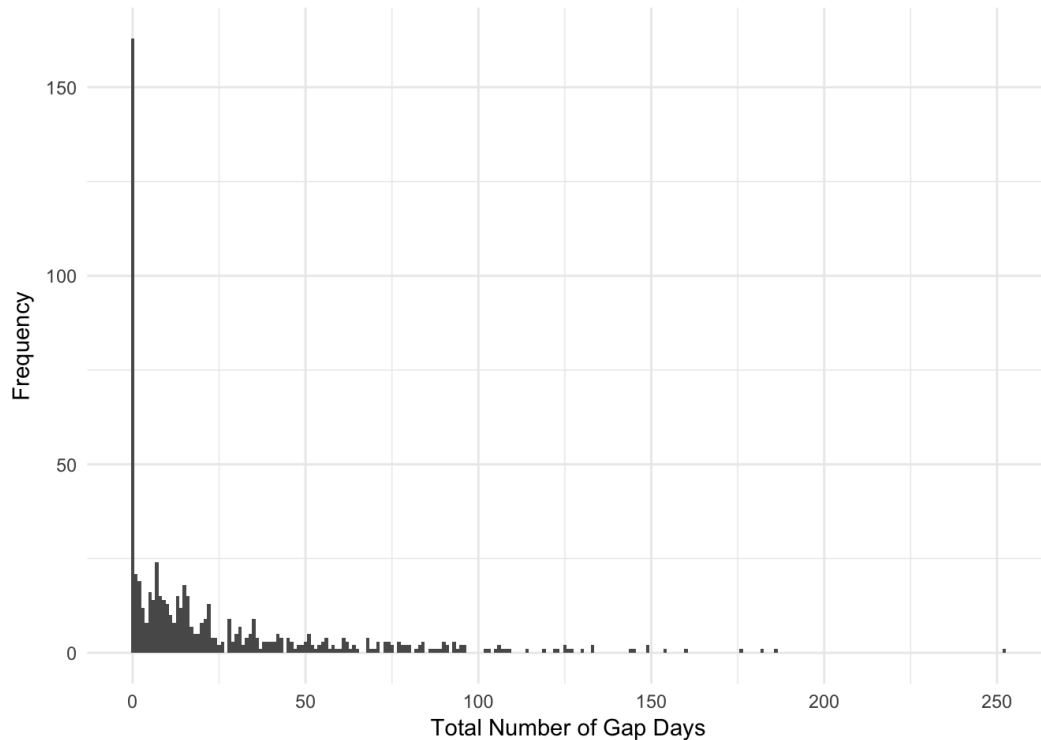


Figure 3.3: Total number of cumulative gap days per patient. Surplus days are not taken into account in the determination of cumulative gap days. The last fill plus method was used to determine the end date for this plot.

long gaps being 130, 124, 119, 107 and 106 days long. Due to the nature of the EHR, it is impossible to know if these long gaps are actually gaps in medication, if the patient has switched to a new pharmacy in the interim, or if medication was discontinued and re-initiated at a later time.

Without this information, it may be of value to consider what constitutes a permissible gap in the setting of the MS data. Permissible gaps need not be determined by gap days and can also be calculated based on interval PDC. Figures 3.4 and 3.5 show the proportion of patients based on various thresholds for a non-permissible gap. Using a threshold to dichotomize permissible versus non-permissible is not recommended unless there have been studies validating that the treatment is effective up to the threshold, but not beyond. Alternatively, this repeated dichotomization could be effective as a means of determining when an intervention should occur.

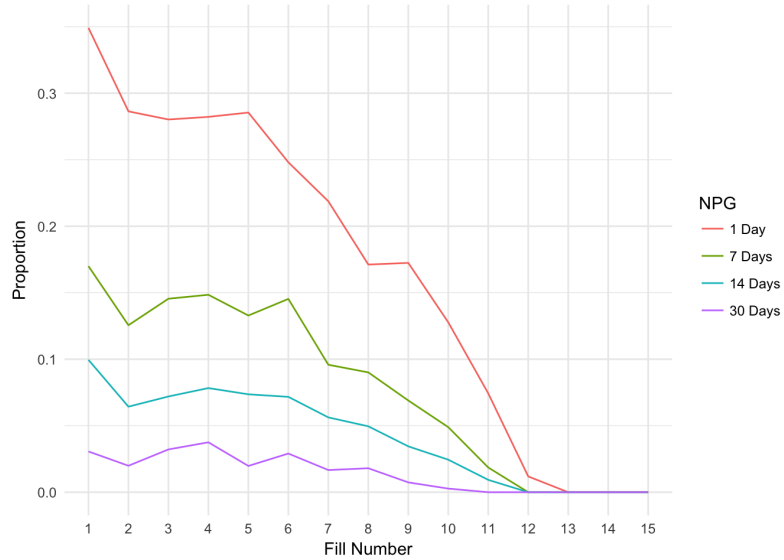


Figure 3.4: Proportion of patients with non-permissible gap by various thresholds for days without medication. At the first fill, almost 35% of patients were at least one day late refilling their prescription, by the eighth fill, this number is approximately reduced by half, and by the thirteenth fill, the remaining patients did not incur any gaps. **NPG**: Non-permissible gap.

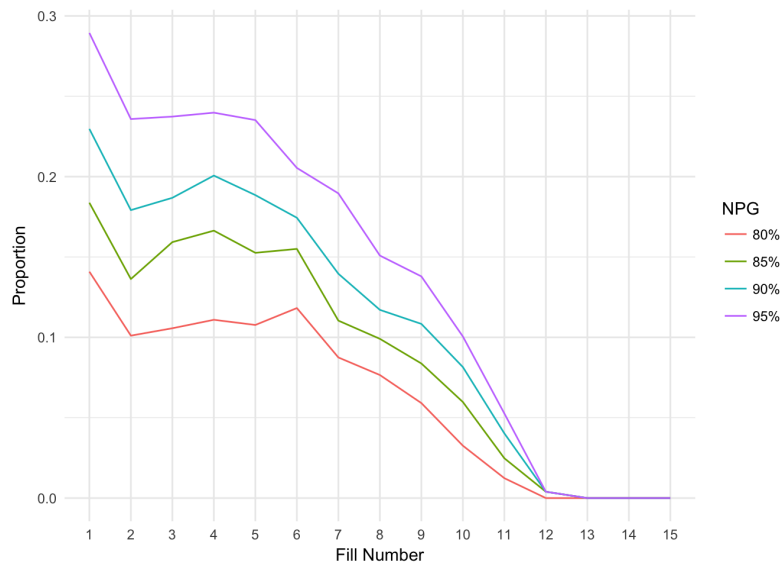


Figure 3.5: Proportion of patients by various thresholds to dichotomize permissible versus non-permissible based on interval PDC. Almost 15% of patients dropped below an interval-based PDC threshold of 80% at the first fill. This rate remained relatively constant until the seventh fill, after which patients' medication adherence patterns improved. **NPG**: Non-permissible gap.



Rather than categorizing the data, we can investigate the timing of each individual fill. Figure 3.6 shows how early or late patients refilled each medication. Because this is calculated as a function of the time of two prescription fills, this calculation does not include the last fill. The highest density in the plot is from minus two to zero days, meaning that the most common timing of a refill is on or within a few days of the expected exhaustion date of the previous fill. There is a large uptick at minus six days, which suggests a structural effect in the data, perhaps something along the lines of an electronic reminder set for one week prior to the medication running out.

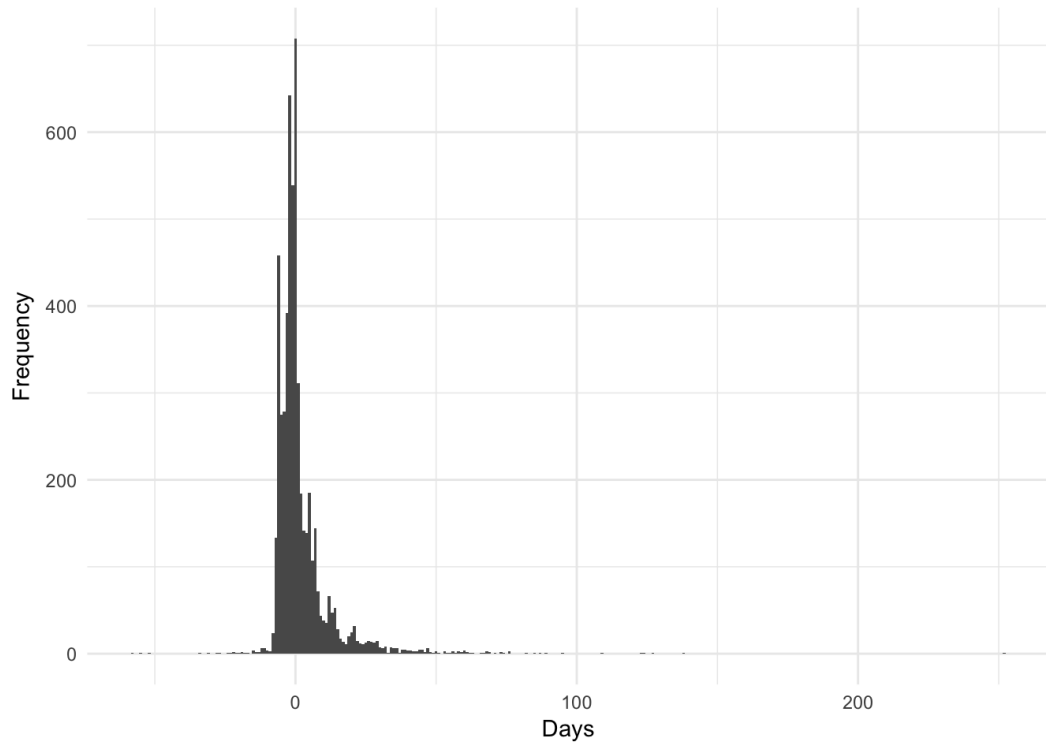


Figure 3.6: Timing of fills at each interval. Positive values represent length of time between when the supply is exhausted and refilling the medication (gap days). Negative values represent the number of days the prescription was filled early (medication surplus).

### 3.2 Sensitivity analysis

In order to determine which of the candidate models will provide the best modeling strategy to medication adherence data, we examine the methods of PDC calculations by

a sensitivity analysis. Sensitivity analyses are crucial in determining how robust methods are to changes in the structure of the data. They aid in the assessment of the robustness of key assumptions, influential observations as well as different modeling methods. The key assumptions being made when calculating PDC are how best to determine the end of the observation period, and how to define the cohort based on complete follow up and the number of fills. We compare different derivations of PDC and patient inclusion criteria. The four PDC derivations are: Fixed, LFP, LFPP, and DBLF. The cohort was restricted by the following rules: follow up through the end of the study period, a minimum of six fills or a minimum of nine fills.

We present the results from sensitivity analyses for four different regression models: logistic, ordinal, negative binomial, and GEE with an identity link. Each of the summary models was fit on the seven different datasets using the same covariates for comparison purposes, while the longitudinal models were fit using only the four different cohort definitions. The covariates include: treatment naive (yes vs. no), financial assistance (copay vs. none), race (non-white vs. white), age (in years), gender (male vs. female), and the drug being prescribed.

For each sensitivity analysis, we compared the estimate of the regression coefficients ( $\beta$ ), the standard error of the estimate (SE) and the associated p-value (p). Depending on the model, the  $\beta$  has a different interpretation. Standard errors are expected to be larger in the models with a smaller sample size. P-values provide some insight into the relative importance of each predictor.

### 3.2.1 Logistic regression

Table 3.4 shows the results from the sensitivity analysis using a logistic regression model. The  $\beta$  coefficient from these models represents the log odds ratio of being more than 80% adherent as defined by PDC. For the Fixed method, we see that the coefficient is 0.42 for copay, which means that a patient receiving copay assistance is expected to have

a log odds ratio of 0.42 of being adherent at the 80% threshold compared to a patient with the same covariate profile, but not receiving copay assistance. Exponentiating this result gives an odds ratio (OR) of 1.52, which means that copay assistance is associated with an 52% increase in being adherent at the 80% threshold, holding all other covariates constant. This predictor is significant in six of the seven models, and has a non-significant p-value in the models which restricts the minimum number of fills to nine. Also, in the nine-fill minimum model, the magnitude of the effect even changes direction. None of the other variables appear to be significant in the prediction of adherence. These results suggest that the logistic model is sensitive to changes in the structure of the data.

Table 3.4: Sensitivity analysis - logistic model

Variable	Stat	By end date				By cohort		
		Fixed (N = 653)	LFP (N = 653)	LFPP (N = 653)	DBLF (N = 653)	Complete (N = 516)	Six fills (N = 516)	Nine fills (N = 406)
Treatment naive	$\beta$	-0.101	0.447	0.126	0.578	0.774	0.390	8.143
	SE	0.237	0.323	0.268	0.320	0.441	0.419	29.416
	p	0.670	0.166	0.638	0.071	0.079	0.351	0.782
Copay vs. none	$\beta$	0.424	0.656	0.528	0.657	0.738	0.530	-0.388
	SE	0.206	0.245	0.222	0.237	0.299	0.302	0.528
	p	0.040	0.008	0.018	0.006	0.013	0.080	0.462
Non-white vs. white	$\beta$	-0.243	-0.249	-0.291	-0.383	-0.028	-0.541	-0.487
	SE	0.246	0.297	0.265	0.281	0.380	0.335	0.500
	p	0.323	0.401	0.271	0.174	0.942	0.106	0.330
Age (per 1 year)	$\beta$	0.015	0.014	0.018	0.008	0.005	-0.002	0.021
	SE	0.009	0.011	0.010	0.011	0.014	0.013	0.020
	p	0.093	0.198	0.072	0.441	0.724	0.872	0.273
Male vs. female	$\beta$	0.048	-0.182	-0.046	-0.226	-0.151	0.023	0.364
	SE	0.221	0.262	0.239	0.251	0.316	0.326	0.520
	p	0.830	0.487	0.849	0.367	0.633	0.943	0.484

**LFP:** last fill plus; **Fixed:** fixed end date; **DBLF:** day before last fill; **LFPP:** last fill plus plus, **Complete:** patients who were followed through to the end of the study period; **Six fills:** patients with at least six fills; **Nine fills:** patients with at least nine fills

### 3.2.2 Ordinal regression

The results of the sensitivity analysis for the ordinal model are shown in Table 3.5. The  $\beta$  coefficient represents the log odds ratio of being more adherent, as defined by having a

higher PDC value. For the LFP method, the log odds ratio for treatment naive compared to patients who are continuing treatment is 0.53, holding all other variables in the model constant. On the odds ratio scale of 1.70, this represents a 70% increase in the odds of having a higher level of adherence for patients who are new to treatment compared to those who are not. While copay assistance was significant in almost every method under the logistic model, it is not a significant predictor in any of the ordinal models. On the other hand, treatment naive is a significant predictor in three of the seven models, and the direction of the effect is constant across models, although the magnitude ranges from 0.214 to 0.620. Age has a very consistent effect and is a significant predictor across the end date rules, but is not significant in the cohort definitions. Thus, the ordinal model can also be considered to be sensitive to changes in the data.

Table 3.5: Sensitivity analysis - ordinal model

Variable	Stat	By end date				By cohort		
		Fixed (N = 653)	LFP (N = 653)	LFPP (N = 653)	DBLF (N = 653)	Complete (N = 516)	Six fills (N = 516)	Nine fills (N = 406)
Treatment naive	$\beta$	0.214	0.526	0.257	0.532	0.620	0.250	0.366
	SE	0.184	0.183	0.181	0.183	0.213	0.225	0.313
	p	0.243	0.004	0.156	0.004	0.004	0.268	0.242
Copay vs. none	$\beta$	0.147	0.183	0.179	0.202	0.123	0.071	-0.136
	SE	0.157	0.158	0.158	0.159	0.179	0.187	0.220
	p	0.350	0.249	0.258	0.203	0.489	0.706	0.536
Non-white vs. white	$\beta$	-0.408	-0.467	-0.463	-0.469	-0.464	-0.504	-0.342
	SE	0.184	0.186	0.185	0.186	0.209	0.214	0.244
	p	0.027	0.012	0.012	0.012	0.026	0.018	0.161
Age (per 1 year)	$\beta$	0.015	0.015	0.015	0.015	0.009	0.008	0.008
	SE	0.007	0.007	0.007	0.007	0.008	0.008	0.009
	p	0.022	0.029	0.024	0.029	0.235	0.318	0.395
Male vs. female	$\beta$	0.073	-0.056	0.054	-0.054	-0.016	0.014	0.084
	SE	0.159	0.160	0.160	0.160	0.180	0.185	0.209
	p	0.647	0.727	0.736	0.734	0.930	0.938	0.688

**LFP**: last fill plus; **Fixed**: fixed end date; **DBLF**: day before last fill; **LFPP**: last fill plus plus, **Complete**: patients who were followed through to the end of the study period; **Six fills**: patients with at least six fills; **Nine fills**: patients with at least nine fills

Table 3.6: Sensitivity analysis - negative binomial model

Variable	Stat	By end date				By cohort		
		Fixed (N = 653)	LFP (N = 653)	LFPP (N = 653)	DBLF (N = 653)	Complete (N = 516)	Six fills (N = 516)	Nine fills (N = 406)
Treatment naive	$\beta$	0.055	-0.263	-0.179	-0.260	-0.390	-0.071	-0.311
	SE	0.152	0.151	0.142	0.152	0.176	0.166	0.210
	p	0.720	0.081	0.207	0.087	0.026	0.669	0.139
Copay vs. none	$\beta$	-0.162	-0.212	-0.222	-0.204	-0.173	-0.186	0.072
	SE	0.125	0.121	0.114	0.122	0.138	0.134	0.154
	p	0.193	0.080	0.052	0.095	0.210	0.165	0.640
Non-white vs. white	$\beta$	0.138	0.226	0.228	0.218	0.147	0.254	0.280
	SE	0.145	0.141	0.133	0.141	0.160	0.151	0.165
	p	0.342	0.108	0.086	0.122	0.358	0.092	0.089
Age (per 1 year)	$\beta$	-0.011	-0.006	-0.007	-0.005	-0.001	0.002	-0.003
	SE	0.005	0.005	0.005	0.005	0.006	0.005	0.006
	p	0.029	0.268	0.149	0.323	0.905	0.713	0.576
Male vs. female	$\beta$	-0.064	0.008	-0.024	0.014	0.089	-0.049	-0.104
	SE	0.126	0.122	0.115	0.122	0.137	0.131	0.144
	p	0.614	0.946	0.833	0.911	0.513	0.712	0.470

**LFP**: last fill plus; **Fixed**: fixed end date; **DBLF**: day before last fill; **LFPP**: last fill plus plus, **Complete**: patients who were followed through to the end of the study period; **Six fills**: patients with at least six fills; **Nine fills**: patients with at least nine fills

### 3.2.3 Negative binomial regression

Table 3.6 shows the sensitivity analysis results using negative binomial regression. Of note is the fact that most of the coefficients are negative, whereas in previous models they were predominantly positive. This is because under the NB framework, we are modeling non-adherence. In this setting,  $\beta$  is the log incidence rate ratio (IRR) of the occurrence of gap days. For the Complete model, treatment naive is a significant predictor with  $\beta = -0.39$ , which means that holding all other variables in the model constant, treatment naive patients have an average IRR of  $\exp(-0.39) = 0.68$  compared to those who are continuing therapy; being new to the medication is associated with a 32% reduction in the number of gap days. Copay assistance is near significance in three of the models, while treatment naive is significant in only one model, but close to significance in two more models. The magnitude of the treatment naive effect is highly variable among the different data structures. Thus, the NB method is also sensitive to the different definitions of the data.

Table 3.7: Sensitivity analysis - GEE model

Variable	Stat	Full (N = 653)	Complete (N = 516)	Six fills (N = 516)	Nine fills (N = 406)
Treatment naive	$\beta$	0.843	1.558	0.347	0.959
	SE	0.582	0.540	0.541	0.462
	p	0.148	0.004	0.521	0.038
Copay vs. none	$\beta$	1.359	-0.066	0.685	-0.212
	SE	0.600	0.518	0.530	0.459
	p	0.023	0.898	0.197	0.643
Non-white vs. white	$\beta$	-1.466	-0.497	-1.329	-0.750
	SE	0.639	0.488	0.563	0.504
	p	0.022	0.308	0.018	0.137
Age (per 1 year)	$\beta$	0.021	0.024	0.017	0.012
	SE	0.021	0.020	0.018	0.016
	p	0.312	0.230	0.343	0.460
Male vs. female	$\beta$	-0.362	-0.266	-0.104	-0.029
	SE	0.529	0.473	0.455	0.390
	p	0.494	0.574	0.819	0.941

The end date rules for all models utilize the day before last fill method; **Full**: all patients included; **Complete**: patients who were followed through to the end of the study period; **Six fills**: patients with at least six fills; **Nine fills**: patients with at least nine fills

### 3.2.4 GEE regression

The sensitivity analysis for the GEE model uses gap and surplus days at each interval as the outcome, and the results are shown in Table 3.7. There are only four datasets in this analysis because the outcomes are generated on the basis of being between two filling events, thus the last fill is not accounted for. The  $\beta$  coefficient for this model is the estimate of the number of days the patient refilled their medication before or after the prescribed date that the previous refill would be used up, where negative values represent a gap in medication. In the DBLF model, race is a significant predictor with  $\beta = -1.47$ . This means that, on average, non-white patients are expected to fill approximately 1.5 days later than white patients. As with the other regression models, the predictors of interest are not consistently significant, or non-significant, and can therefore be considered sensitive to alterations in the data.

### 3.2.5 Model comparison

We compared the relative performance between models, and the results are in Table 3.8, which also includes results from two different logistic GEE models, one using the occurrence of any gap time as the outcome, and the other using an interval PDC of 80%. Each regression model was fit using the day before last fill rule, so the summary and longitudinal models can be comparable using the same data.

Treatment naive is significant, or nearly significant in the summary models, but not with the longitudinal models. Copay assistance is significant, or near significant, in every model, suggesting that it is indeed an important predictor of medication adherence. Race is an important predictor in many of the models as well. Both sex and age are consistently non-significant across models.

Table 3.8: Method comparison

Variable	Stat	Summary models			Longitudinal models		
		NB	Logistic	Ordinal	GEE(C)	GEE(B1)	GEE(B80)
Treatment naive	$\beta$	-0.260	0.578	0.532	0.843	-0.117	-0.006
	SE	0.152	0.320	0.183	0.582	0.137	0.017
	p	0.087	0.071	0.004	0.148	0.393	0.731
Copay vs. none	$\beta$	-0.204	0.657	0.202	1.359	-0.259	-0.029
	SE	0.122	0.237	0.159	0.600	0.117	0.017
	p	0.095	0.006	0.203	0.023	0.027	0.081
Non-white vs. white	$\beta$	0.218	-0.383	-0.469	-1.466	0.305	0.044
	SE	0.141	0.281	0.186	0.639	0.135	0.020
	p	0.122	0.174	0.012	0.022	0.024	0.030
Age (per 1 year)	$\beta$	-0.005	0.008	0.015	0.021	0.003	0.000
	SE	0.005	0.011	0.007	0.021	0.005	0.001
	p	0.323	0.441	0.029	0.312	0.580	0.677
Male vs. female	$\beta$	0.014	-0.226	-0.054	-0.362	0.055	0.009
	SE	0.122	0.251	0.160	0.529	0.117	0.016
	p	0.911	0.367	0.734	0.494	0.640	0.575

The end date rules for all models utilize the day before last fill method and the full cohort; (C): GEE using gap and surplus days as outcome; (B1): binomial GEE using a 1 day non permissible gap; (B80): binomial GEE using a 80% interval PDC non permissible gap

### 3.3 Model checking and comparison

Based on the results from the sensitivity analyses and the method comparison, it is not clear which model is the best to fit the data. This is, in part, due to the fact that it is difficult to discern what the truth is when the truth is unknown. Another method to check the performance of the model is to compare how the predicted values from each model fit compare to the observed data. However, different types of models (e.g., logistic model vs. NB vs. GEE) provide different predictions. The logistic model can only estimate the probability of being adherent based on a threshold of 80% PDC. We can calculate exceedence probabilities using the ordinal model and compare those two models against one another as well as the observed data. The ordinal model can also estimate the mean PDC, as well as quantiles - in our case, we will examine the estimated median PDC. We can use the predicted number gap days from the NB model and the average length of the observation window to calculate PDC. Similarly, we can use the predicted timing of refills from the GEE model, the average number of fills and length of the observation window to calculate PDC. Thus, we can compare the estimated PDCs for the NB, ordinal and GEE models against the observed values. However, because the NB and GEE models are not designed to predict PDC, we will also compare their model-specific predictions against observed values.

For the model checking, a reduced model was fit on the MS data due to low frequencies in some of the cells of covariates. The three predictors that appear to be the most important for predicting adherence based on the sensitivity analysis were chosen: treatment naive, copay assistance and race. To be able to compare GEE to the summary models, the DBLF definition was used for all models.



### 3.3.1 Predicted PDC

Table 3.9 shows the observed and estimated PDC for the GEE, NB and ordinal models across eight levels of covariate combinations. There is no clear winner among the three models, however, the ordinal model provides the smaller difference. In general, the ordinal model under-predicted the PDC, and the two instances of the largest difference are in cells with a low frequency. The NB model also performs poorly for the cells with a low frequency. Neither the NB model nor the GEE model directly predict PDC, and represent a second- or third-level approximation based on the denominator and fill data.

### 3.3.2 Predicted PDC $\geq$ 80%

While we do not recommend dichotomizing PDC as the outcome, the results in Table 3.10 are included for illustrative purposes. Neither model consistently predicts adherence at the 80% threshold. For both models, very large differences were observed in low frequency cells. The logistic model appears to outperform the ordinal model, though not by much.

Table 3.9: Model checking and comparison - predicted mean PDC

Naive	Copay	Race	GEE				Ordinal				NB	
			N	Observed	Predicted	% Difference	Predicted	% Difference	Predicted	% Difference	Predicted	% Difference
No	White		109	0.893	0.908	1.732	0.881	-1.303	0.892	-0.122		
	Non-White		27	0.827	0.884	6.972	0.848	2.588	0.863	4.344		
Yes	White		328	0.921	0.943	2.348	0.898	-2.563	0.917	-0.466		
	Non-White		54	0.899	0.901	0.137	0.867	-3.637	0.900	0.117		
No	White		35	0.928	0.927	-0.115	0.910	-1.898	0.841	-9.405		
	Non-White		7	0.930	0.887	-4.639	0.881	-5.210	0.851	-8.489		
Yes	White		75	0.925	0.969	4.838	0.925	0.011	0.903	-2.351		
	Non-White		18	0.940	0.915	-2.601	0.898	-4.456	0.895	-4.805		

Table 3.10: Model checking and comparison - probability of PDC  $\geq 80\%$

Naive	Copay	Race	N	Observed	Logistic				Ordinal	
					Predicted	% Difference	Predicted	% Difference	Predicted	% Difference
No	White		109	0.798	0.774	-3.070	0.738	-7.49		
	Non-White		27	0.593	0.701	18.363	0.822	38.69		
Yes	White		328	0.869	0.870	0.131	0.822	-5.38		
	Non-White		54	0.833	0.821	-1.425	0.883	5.98		
No	White		35	0.829	0.859	3.720	0.787	-5.00		
	Non-White		7	1.000	0.808	-19.229	0.858	-14.20		
Yes	White		75	0.907	0.923	1.791	0.858	-5.33		
	Non-White		18	0.944	0.892	-5.593	0.908	-3.83		

### 3.3.3 Predicted gap days

The NB model is better suited to predicting gap days rather than PDC, and the predicted gap days are shown in Table 3.11. The difference is very large in many of the cells, however, larger frequencies provide smaller differences. The model over-predicts the number of gap days in all but one instance.

Table 3.11: Model checking and comparison - predicted gap days

Naive	Copay	Race	N	Observed	NB	
					Predicted	% Difference
No	No	White	109	28.5	30.6	7.33
		Non-White	27	48.6	37.8	-22.29
	Yes	White	328	23.7	24.7	4.32
		Non-White	54	28.8	30.5	6.00
Yes	No	White	35	14.9	24.3	63.03
		Non-White	7	16.9	30.1	78.26
	Yes	White	75	17.4	19.6	13.14
		Non-White	18	14.6	24.3	66.76

### 3.3.4 Predicted gap and surplus days

The GEE methodology assumes a different structure for the outcome compared to the summary models, and the results of predicting the timing of refills are shown in Table 3.12. There are large differences between the predicted and observed values. This may be partially attributable to the skewed distribution of refill time (see Figure 3.6). Overall, this method does not appear to provide a good fit to the MS data.

### 3.3.5 Predicted median PDC

Lastly, we looked at the ability of the ordinal model to predict the median PDC (Table 3.13). Very large differences are not observed across the predicted values. The versatility of the ordinal model appears to make it a good candidate model for fitting adherence data.

Table 3.12: Model checking and comparison - predicted gap and surplus days

Naive	Copay	Race	N	Observed	GEE	
					Predicted	% Difference
No	No	White	109	-4.151	-3.236	-22.1
		Non-White	27	-8.416	-4.742	-43.7
	Yes	White	328	-2.234	-1.809	-19.0
		Non-White	54	-3.706	-3.315	-10.6
Yes	No	White	35	-1.516	-2.393	57.8
		Non-White	7	-2.364	-3.899	65.0
	Yes	White	75	-1.813	-0.966	-46.7
		Non-White	18	-0.836	-2.472	195.9

Table 3.13: Model checking and comparison - predicted median PDC

Naive	Copay	Race	N	Observed	Ordinal	
					Predicted	% Difference
No	No	White	109	0.954	0.917	-3.862
		Non-White	27	0.897	0.952	6.199
	Yes	White	328	0.962	0.952	-1.015
		Non-White	54	0.938	0.973	3.690
Yes	No	White	35	0.983	0.939	-4.530
		Non-White	7	0.948	0.963	1.589
	Yes	White	75	0.968	0.963	-0.490
		Non-White	18	0.984	0.979	-0.502

## Chapter 4

### Simulation study

We will perform Monte Carlo simulation to generate mock adherence datasets with known conditions. We will then apply each of the different cohort definitions on the simulated datasets, fit each dataset using the regression models discussed in the previous chapters, and compare the performance of each against the known truth.

#### 4.1 Model specifications

Since PDC is a composite outcome generated from two quantities (i.e., the numerator and denominator), instead of simulating PDC, we mimic the actual data-generating process by simulating the timing of each refill. Through a process of trial and error, we were able to generate adherence data that is comparable to what was observed in the MS study.

The first thing to notice is that the distribution of the gap and surplus days at each fill is skewed in such a way that there are more gap days than surplus days, and that the longest gaps exceed the length of the surplus days (Figure 3.3). This is a structural component of the data in that the amount of surplus can not exceed the amount of drug supplied at the previous fill - if it did the current fill in the sequence would predate the previous fill. To approximate this distribution, we applied a non-central t-distribution-normal random-effects model as follows:

For  $i = 1, \dots, N, j = 1, \dots, m_i,$

$$Y_{ij} \sim T_v(\delta_i | X_i, m_i,)$$

$$\delta_i = \beta_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i},$$

$$\beta_{0i} \sim N(0, \sigma^2),$$

where  $Y_{ij}$  is the refill time for patient  $i$  at fill number  $j$ , following a non-central t-distribution,  $T_\nu(\delta)$ .  $\delta_i$  is the non-centrality parameter for patient  $i$ ,  $\nu$  is the degrees of freedom set at 2.3,  $X_i = (X_{1i}, X_{2i})$  is the covariate matrix,  $\beta_{0i}$  is the patient-specific random effect, with the variance  $\sigma^2$  fixed at 12. We used a non-central t-distribution with  $\nu = 2.3$  to mimic the observed gaps and surplus days as close as possible to the MS data. The non-centrality parameter determines the skewness of the distribution, and larger  $\delta$  increases the likelihood of a larger gap between fills.

Each patient is assumed to have his or her own propensity for the timing of their refills, that is, some patients will commonly refill early, others late, and others mostly on time, barring special circumstances. This subject-specific propensity for filling is modeled as a random effect following a normal distribution with mean  $\beta_1 X_{1i} + \beta_2 X_{2i}$ . We considered two binary covariates;  $X_1$  is a real predictor on adherence with  $\beta = 2$ , whereas  $X_2$  is not associated with the propensity for adherence, thus  $\beta_2 = 0$ .

In order to further mimic the MS data, some additional modifications were necessary. Because the majority of pills dispensed at each refill was either 28 or 30 days (Table 3.2), all patients were assigned 30 days worth of supply at each fill. To address the possibility for improbably early fills, any large surplus was truncated at 15 days. This truncation is likely a source of bias in the model, though we found that it was between one and two percent of all fill events from the simulation results.

Not all patients had the same number of fills, nor did they have the capacity to achieve similar numbers of fills, as patients were enrolled into the study at different times. In the MS study, approximately 79% of patients were continuing therapy, and thus were observed for almost the full year. The remainder of patients were treatment naive, and were enrolled continuously through the year up until October (Figure 4.1). To accommodate this feature, patients were assigned a start date based on two possible random distributions. Continuing patients were given a start data based on an Exponential distribution:  $t_{0C} \sim Exp(0.075)$ .

This made the start date for most of these patients within the first 30 days of the study. Treatment naive patients were assigned a start date based on a Discrete Uniform distribution:  $t_{0N} \sim U(15, 275)$ , so that these patients were randomly enrolled throughout the first nine months of the study period.

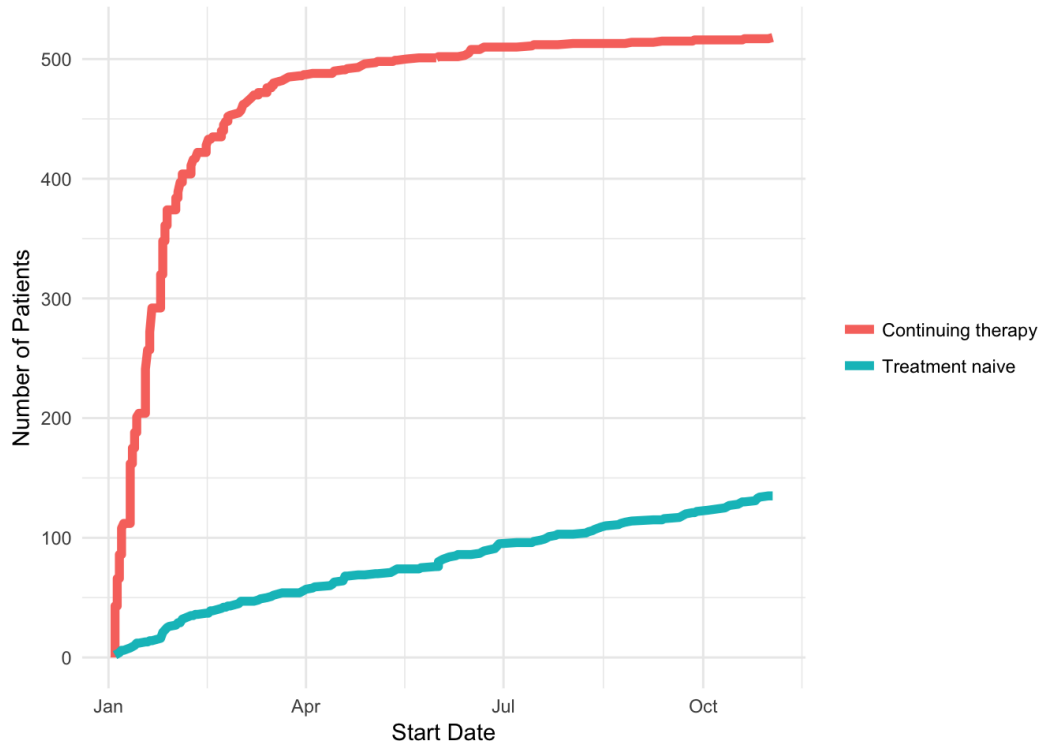


Figure 4.1: Cumulative number of enrolled patients by treatment status. The majority (79%) of patients were continuing therapy from before the study start date. The red line shows that most of those patients refilled their medication within the first month of the study, as would be typical for a month-supply of medication. Large gaps coinciding with the start of the study period could explain the delay for some patients whose first fills were not until the middle of the year. For treatment naive patients, enrollment happened continuously throughout the first ten months of the study, at which point, attaining the required minimum of three fills is no longer achievable.

Lastly, there were 116 patients with a gap of longer than 30 days, and 71 patients whose last fill occurred 90 days prior to the end of the study period in the MS data. While the distribution is skewed, it does not provide the extreme level of skewness to mimic this result. To approximate this high skewness, 120 fill events were randomly assigned a long gap between 30 and 450 days. The end of the study period was set at 365 days, and thus any

gap extending beyond this date is considered a drop out in the context of the simulation.

Under these simulation conditions, 5000 replicated datasets were generated with a total of 670 subjects for each replication. From each simulated dataset, the outcome was defined using the day before last fill rule. To account for inclusion criteria that subjects have at least three fills, subjects with fewer than three fills were excluded from each replicated data. Four types of inclusion criteria were applied to the data, which are: the full cohort ( $\geq 3$  fills), subjects with follow up through the end of the end of the observation period, and subjects with a minimum of either six or nine fills. Four models (logistic, ordinal, NB and GEE) were fit on each of the four different data variants, yielding a total of sixteen model fits for each simulation. Variables of interest were extracted from the results of each of the models and stored in a data frame at the end of each iteration.

## 4.2 Simulation Results

Overall characteristics of simulated datasets by cohort definition are presented in Table 4.1. The average PDC is around 91% for the four different cohorts, which is quite similar to what was seen in the MS data. Standard deviations (SDs) and medians are also comparable to the MS data, but the percentage of patients below 80% PDC is slightly different from the MS data. In the full dataset, only 1.33% of the filling events were truncated due to a fill that was deemed “improbably early.”

Table 4.1: Simulation Results - Average Cohort Statistics

Data	N	PDC			$\leq 80\%$ PDC		
		Mean	SD	Median	N	%	% Truncated
Full	663	0.906	0.118	0.947	105.8	16.0	1.33
Complete follow up	530	0.913	0.117	0.960	76.5	14.4	1.11
Six fills	602	0.913	0.106	0.952	90.0	14.9	1.40
Nine fills	510	0.926	0.086	0.962	57.5	11.3	1.50

**Full:** the full cohort; **Complete:** patients with follow up up through the end of the study period; **Six fills:** patients with at least six fills; **Nine fills:** patients with at least 9 fills.



#### 4.2.1 Power and type I error

We calculate the power and type I error rate (at the 0.05 level) of each model (Table 4.2). Power is the ability of a test to detect a departure from the null hypothesis when the null hypothesis is false, while type I error is the probability that the null hypothesis is rejected, when the null is true. In our setting,  $\beta_1$  has a true effect on the timing of refills. The ordinal model performs the best, with nearly 100% power for all four cohort variations. The logistic model does the worst, with a lowest power of 86% for the cohort restricted to a minimum of nine fills. This is not surprising, given that dichotomization is known to reduce power. On the other hand,  $\beta_2$  was specified to have a null effect on patient adherence. The logistic model has a type I error rate below 5% for three of the cohort derivations. For both the ordinal and GEE models, the type I error rate is just above 5%, while the NB model has a type I error rate of almost 7% for the full and complete follow-up cohorts. Overall, it appears that the ordinal model performs the best, however, the type I error rate is slightly inflated for most models.

Table 4.2: Simulation Results - Power and Type I Error

Cohort	$H_0 : \beta = 0$	NB	Logistic	Ordinal	GEE
Full	Power	99.38	97.02	100.00	99.94
	Type I error	6.94	4.90	5.28	5.20
Complete follow up	Power	95.92	88.10	99.96	98.74
	Type I error	6.98	4.70	5.30	5.44
Nine fills	Power	99.82	86.42	99.98	100.00
	Type I error	4.70	4.38	4.96	4.90
Six fills	Power	99.88	95.74	100.00	99.98
	Type I error	5.84	5.12	5.34	5.48

#### 4.2.2 Predicted PDC

Only the ordinal model can directly predict PDC, however, just as in Section 3.3.1, we can use gap days, the length of the observation window, and the number of fills to calculate

PDC for the NB and GEE models (Table 4.3). The GEE model shows a large bias for the models in which  $X_1 = 0$ . This is because the GEE model is taking surplus days into account, and subjects without  $X_1$  will incur more surplus days, thus estimates of PDC are higher. When  $X_1 = 1$ , the bias for predicted PDC is as much as 2%. The ordinal model has consistently smaller bias, but it underestimates the observed value of PDC. The NB model performs the best in predicting PDC under the simulation conditions, with a largest bias of just 0.25%.

Table 4.3: Simulation Results - Predicted Mean PDC

Cohort	Covariates ( $X$ )	Observed	GEE		Ordinal		NB	
			Predicted	% Difference	Predicted	% Difference	Predicted	% Difference
Full	$X_1 = 0, X_2 = 0$	0.929	0.976	5.097	0.922	-0.756	0.931	0.254
	$X_1 = 1, X_2 = 0$	0.883	0.889	0.673	0.874	-1.094	0.881	-0.217
	$X_1 = 0, X_2 = 1$	0.929	0.976	5.106	0.922	-0.753	0.931	0.251
	$X_1 = 1, X_2 = 1$	0.883	0.889	0.688	0.874	-1.080	0.881	-0.207
Complete follow up	$X_1 = 0, X_2 = 0$	0.933	0.980	5.035	0.927	-0.640	0.935	0.242
	$X_1 = 1, X_2 = 0$	0.892	0.903	1.168	0.884	-0.925	0.890	-0.232
	$X_1 = 0, X_2 = 1$	0.933	0.980	5.033	0.927	-0.642	0.935	0.237
	$X_1 = 1, X_2 = 1$	0.892	0.903	1.193	0.884	-0.900	0.890	-0.206
Nine fills	$X_1 = 0, X_2 = 0$	0.945	1.006	6.432	0.942	-0.256	0.945	0.013
	$X_1 = 1, X_2 = 0$	0.906	0.924	2.015	0.904	-0.262	0.906	-0.011
	$X_1 = 0, X_2 = 1$	0.945	1.006	6.468	0.942	-0.247	0.945	0.028
	$X_1 = 1, X_2 = 1$	0.906	0.924	2.024	0.904	-0.277	0.906	-0.016
Six fills	$X_1 = 0, X_2 = 0$	0.935	0.988	5.681	0.930	-0.500	0.936	0.116
	$X_1 = 1, X_2 = 0$	0.890	0.900	1.175	0.885	-0.600	0.889	-0.095
	$X_1 = 0, X_2 = 1$	0.935	0.988	5.712	0.930	-0.494	0.936	0.127
	$X_1 = 1, X_2 = 1$	0.890	0.901	1.186	0.885	-0.609	0.889	-0.099

### 4.2.3 Predicted PDC $\geq 80\%$

In predicting the probability of PDC  $\geq 80\%$ , the logistic model resulted in little bias. The ordinal model performed well too, with a maximum percent bias of 0.5%. This supports that the ordinal model should be preferred over the logistic model as the former can not only predict PDC  $\geq 80\%$  almost as well as the logistic model, but also can provide much more useful information such as estimates of quantiles and the mean.

Table 4.4: Simulation Results - Probability of PDC  $\geq 80\%$

Cohort	Covariates (X)	Observed	Logistic		Ordinal	
			Predicted	% Difference	Predicted	% Difference
Full	$X_1 = 0, X_2 = 0$	0.893	0.893	0.023	0.897	0.483
	$X_1 = 1, X_2 = 0$	0.784	0.784	0.039	0.785	0.137
	$X_1 = 0, X_2 = 1$	0.893	0.893	0.025	0.897	0.489
	$X_1 = 1, X_2 = 1$	0.784	0.784	0.032	0.785	0.132
Complete follow up	$X_1 = 0, X_2 = 0$	0.901	0.902	0.034	0.904	0.354
	$X_1 = 1, X_2 = 0$	0.806	0.806	0.032	0.806	0.096
	$X_1 = 0, X_2 = 1$	0.902	0.901	0.016	0.904	0.356
	$X_1 = 1, X_2 = 1$	0.806	0.806	0.044	0.806	0.147
Nine fills	$X_1 = 0, X_2 = 0$	0.928	0.928	0.001	0.928	0.025
	$X_1 = 1, X_2 = 0$	0.842	0.842	0.054	0.844	0.313
	$X_1 = 0, X_2 = 1$	0.928	0.928	0.038	0.928	0.061
	$X_1 = 1, X_2 = 1$	0.843	0.843	0.008	0.844	0.233
Six fills	$X_1 = 0, X_2 = 0$	0.903	0.903	0.015	0.904	0.244
	$X_1 = 1, X_2 = 0$	0.796	0.796	0.049	0.796	0.168
	$X_1 = 0, X_2 = 1$	0.903	0.903	0.033	0.905	0.272
	$X_1 = 1, X_2 = 1$	0.796	0.796	0.022	0.797	0.140

### 4.2.4 Predicted gap days

The NB model can directly predict the number of gap days that represent non-adherence, which makes the NB model different from other models we discussed. Table 4.5 shows the predicted gap days from the simulation study. The bias is positive for all predictions, ranging from 1% to 8%. Compared to its prediction of PDC, the NB model is apparently not as good at predicting gap days as it is at predicting PDC, but this is not the case. The difference is due to the denominator in the % Difference calculation: the large denomina-

tor in the PDC calculation means that the ratio will be smaller. Thus, it is the same bias in both predictions, just more pronounced when non-adherence rather than adherence is considered.

Table 4.5: Simulation Results - Estimation of Gap Days

Cohort	Covariates (X)	Observed	NB	
			Predicted	% Difference
Full	$X_1 = 0, X_2 = 0$	21.3	23.0	8.02
	$X_1 = 1, X_2 = 0$	34.7	37.1	7.31
	$X_1 = 0, X_2 = 1$	21.3	22.9	8.04
	$X_1 = 1, X_2 = 1$	34.7	37.1	7.25
Complete follow up	$X_1 = 0, X_2 = 0$	20.7	22.0	7.16
	$X_1 = 1, X_2 = 0$	33.1	35.2	6.61
	$X_1 = 0, X_2 = 1$	20.6	22.0	7.23
	$X_1 = 1, X_2 = 1$	33.1	35.1	6.49
Nine fills	$X_1 = 0, X_2 = 0$	18.3	18.7	2.91
	$X_1 = 1, X_2 = 0$	31.2	31.7	1.54
	$X_1 = 0, X_2 = 1$	18.3	18.7	2.75
	$X_1 = 1, X_2 = 1$	31.2	31.6	1.70
Six fills	$X_1 = 0, X_2 = 0$	20.7	21.6	4.88
	$X_1 = 1, X_2 = 0$	34.9	36.0	3.48
	$X_1 = 0, X_2 = 1$	20.7	21.6	4.77
	$X_1 = 1, X_2 = 1$	34.8	36.0	3.58

#### 4.2.5 Predicted median PDC

Table 4.6 shows the simulation results for predicting median PDC using the ordinal model. The percent bias is negligible here, supporting another useful feature of the ordinal model.

#### 4.2.6 GEE model performance

Lastly, we examine how well the GEE model can estimate the true parameters of the simulation. Table 4.7 shows the estimates of  $\beta$ ,  $\hat{\beta}$ , compared to the known truth. Note

Table 4.6: Simulation Results - Estimation of Median PDC

Cohort	Covariates (X)	Observed	Ordinal	
			Predicted	% Difference
Full	$X_1 = 0, X_2 = 0$	0.985	0.984	-0.027
	$X_1 = 1, X_2 = 0$	0.907	0.906	-0.075
	$X_1 = 0, X_2 = 1$	0.985	0.984	-0.021
	$X_1 = 1, X_2 = 1$	0.907	0.906	-0.081
Complete follow up	$X_1 = 0, X_2 = 0$	0.989	0.989	0.003
	$X_1 = 1, X_2 = 0$	0.922	0.921	-0.045
	$X_1 = 0, X_2 = 1$	0.989	0.989	-0.001
	$X_1 = 1, X_2 = 1$	0.922	0.921	-0.018
Nine fills	$X_1 = 0, X_2 = 0$	0.992	0.992	0.009
	$X_1 = 1, X_2 = 0$	0.923	0.922	-0.134
	$X_1 = 0, X_2 = 1$	0.992	0.992	0.005
	$X_1 = 1, X_2 = 1$	0.923	0.922	-0.122
Six fills	$X_1 = 0, X_2 = 0$	0.988	0.988	0.005
	$X_1 = 1, X_2 = 0$	0.911	0.910	-0.102
	$X_1 = 0, X_2 = 1$	0.988	0.988	0.011
	$X_1 = 1, X_2 = 1$	0.911	0.910	-0.106

Table 4.7: Simulation Results - GEE Model Performance

Cohort	Truth	$\hat{\beta}$		$SE(\hat{\beta})$		
		Mean( $\hat{\beta}$ )	Bias	$SD(\hat{\beta})$	$\overline{SE}(\hat{\beta})$	$\frac{SD(\hat{\beta})}{\overline{SE}(\hat{\beta})}$
Full	$\beta_1 = 2$	2.913	0.913	0.568	0.565	1.005
	$\beta_2 = 0$	-0.005	-0.005	0.566	0.565	1.002
Complete follow up	$\beta_1 = 2$	2.553	0.553	0.608	0.602	1.011
	$\beta_2 = 0$	-0.003	-0.003	0.605	0.601	1.006
Nine fills	$\beta_1 = 2$	2.573	0.573	0.434	0.435	0.998
	$\beta_2 = 0$	-0.007	-0.007	0.436	0.435	1.002
Six fills	$\beta_1 = 2$	2.877	0.877	0.491	0.484	1.014
	$\beta_2 = 0$	-0.008	-0.008	0.487	0.484	1.007

Bias =  $\hat{\beta} - \beta$

that absolute bias is reported. There is very large positive bias in estimating  $\beta_1$ . The delay in the timing of refills is overestimated by almost a full day in the full cohort. The

GEE methodology assumes a symmetric distribution, while the data were simulated from a skewed distribution, which is heavily skewed. Large added gaps will skew the estimate of the mean further still. Unfortunately, in this setting the GEE model was not able to overcome these challenges.

Because of the multiple random components of the model, as well as the truncation of surplus time and addition of long gap periods, the true variability of the model is unknown, which cannot be compared with the simulated results. Instead, the SD of the estimated coefficients to the mean of SE of  $\hat{\beta}$ . Overall the bias in the estimate suggests that this method is not well suited to estimating adherence data.

### 4.3 Ordinal model simulations

The ordinal model showed the best overall performance in the above simulation. We observed no major differences within the ordinal model among the different cohort-inclusion rules. Next, we want to evaluate the performance of the ordinal model using the different study end date rules. We perform two additional simulations: one with the same conditions as described above, and another with a smaller sample size and conditions leading to lower adherence.

#### 4.3.1 Under similar adherence conditions

Tables 4.8 and 4.9 show the results from the simulation using the four different end date rules. With 5000 replications, we see no major difference among the different end date rules in regards to power and the type I error rate. Likewise, the percent bias in predicting PDC is consistently small across the four methods. The fixed end date rules shows larger bias than the other methods.

Table 4.8: Ordinal Model Simulation Results - Power and Type I Error

Rule	$H_0 : \beta = 0$	Ordinal
LFP	Power	100.00
	Type I error	5.28
DBLF	Power	100.00
	Type I error	5.26
LFPP	Power	100.00
	Type I error	5.58
Fixed	Power	99.98
	Type I error	5.38

Table 4.9: Ordinal Model Simulation Results - Predicted Mean PDC

Rule	Covariates (X)	Observed	Ordinal	
			Predicted	% Difference
LFP	$X_1 = 0, X_2 = 0$	0.933	0.936	0.249
	$X_1 = 1, X_2 = 0$	0.891	0.889	-0.207
	$X_1 = 0, X_2 = 1$	0.933	0.936	0.246
	$X_1 = 1, X_2 = 1$	0.891	0.889	-0.209
DBLF	$X_1 = 0, X_2 = 0$	0.929	0.931	0.253
	$X_1 = 1, X_2 = 0$	0.883	0.881	-0.211
	$X_1 = 0, X_2 = 1$	0.929	0.931	0.251
	$X_1 = 1, X_2 = 1$	0.883	0.881	-0.213
LFPP	$X_1 = 0, X_2 = 0$	0.927	0.929	0.245
	$X_1 = 1, X_2 = 0$	0.883	0.881	-0.195
	$X_1 = 0, X_2 = 1$	0.927	0.929	0.246
	$X_1 = 1, X_2 = 1$	0.883	0.881	-0.202
Fixed	$X_1 = 0, X_2 = 0$	0.913	0.919	0.620
	$X_1 = 1, X_2 = 0$	0.871	0.867	-0.495
	$X_1 = 0, X_2 = 1$	0.913	0.919	0.615
	$X_1 = 1, X_2 = 1$	0.871	0.867	-0.515

#### 4.3.2 Under less favorable adherence conditions

To simulate data under less favorable conditions, we reduced the sample size to 200 subjects, increased the effect of  $\beta_1$  from 2 to 2.5, reduced the maximum number of refills a subject could have to 12 (down from 16), and increased the proportion of treatment naive



patients to 50%. These alterations is to reduce the overall average PDC, as well as to reduce the number of patients with complete follow up. Under these conditions, the decision on how best to define the study end date has a much larger impact on adherence rates.

In Table 4.10, we present the type I error rate and power for each of the end date derivations. In every case, the power is lower than in the previous simulations. Note that the power is reduced to 70% for the fixed end date rule with the type I error rate being the highest, 5.14%. The LFPP method also has reduced power. Both the LFP and DBLF methods have similar results for type I error and power.

The results for predicting the mean PDC are given in Table 4.11. The first thing to note is how variable the mean PDC is across each of the four end date rules. The mean PDC for fixed end date is by far the smallest, followed by LFPP, DBLF and then LFP. The LFPP method has the smallest overall percent bias, and the fixed end date has the highest. There is no obvious best method, but it is clear from these results that the fixed end date rule is not preferred.

Table 4.10: Low Adherence Simulation Results - Power and Type I Error

Rule	$H_0 : \beta = 0$	Ordinal
LFP	Power	93.70
	Type I error	4.90
DBLF	Power	93.82
	Type I error	4.92
LFPP	Power	89.98
	Type I error	4.98
Fixed	Power	69.08
	Type I error	5.14

Table 4.11: Low Adherence Simulation Results - Predicted Mean PDC

Rule	Covariates (X)	Observed	Ordinal	
			Predicted	% Difference
LFP	$X_1 = 0, X_2 = 0$	0.906	0.915	1.018
	$X_1 = 1, X_2 = 0$	0.858	0.851	-0.765
	$X_1 = 0, X_2 = 1$	0.906	0.915	0.999
	$X_1 = 1, X_2 = 1$	0.858	0.851	-0.750
DBLF	$X_1 = 0, X_2 = 0$	0.899	0.908	1.049
	$X_1 = 1, X_2 = 0$	0.845	0.838	-0.776
	$X_1 = 0, X_2 = 1$	0.899	0.908	1.027
	$X_1 = 1, X_2 = 1$	0.845	0.838	-0.761
LFPP	$X_1 = 0, X_2 = 0$	0.885	0.893	0.852
	$X_1 = 1, X_2 = 0$	0.834	0.829	-0.595
	$X_1 = 0, X_2 = 1$	0.886	0.893	0.827
	$X_1 = 1, X_2 = 1$	0.834	0.829	-0.579
Fixed	$X_1 = 0, X_2 = 0$	0.839	0.851	1.494
	$X_1 = 1, X_2 = 0$	0.796	0.787	-1.043
	$X_1 = 0, X_2 = 1$	0.839	0.851	1.495
	$X_1 = 1, X_2 = 1$	0.796	0.787	-1.046

## Chapter 5

### Discussion

In this thesis, we presented a summary of commonly used adherence measures, discussed the strengths and weaknesses of each, and suggested a framework for the generalization of outcome derivations. We then discussed modeling strategies using common derivations of medication adherence such as PDC as the outcome. Four different statistical methods were considered: logistic, ordinal, negative binomial regressions, and GEE methodology. We presented a case study using medication adherence data from 653 patients and conducted a sensitivity analysis across the four models and seven different outcome generating mechanisms. Finally, we carried out simulation studies to evaluate the performance of the four models using different criteria for patient inclusion.

Ordinal logistic regression performs the best in modeling adherence data, providing reliable estimates of the mean and median PDC values, as well as predicting the  $PDC \geq 80\%$  with comparable accuracy to that of the logistic model. We hope that this thesis can provide a stepping stone towards abandoning the dichotomization of adherence data in future studies. The longitudinal analyses proposed in this analysis did not effectively predict the outcome.

We found no clear evidence that there is any benefit to further restricting the cohort based on the minimum number of fills, or by limiting the cohort only to patients with complete follow up, hence we promote using the full cohort ( $\geq 3$  fills) as the best practice. Additionally, we prefer the DBLF rule of determining the outcome. This method performed as well as, or better than the other methods. Finally, DBLF is the only method which is compatible with longitudinal data analyses, should superior alternatives be found.

While the GEE model was not as good as the summary models, the work required to be able to apply them presents a more accurate profile of medication adherence. Looking at

individual refill intervals allows for a more fair comparison between patients, and provides a better understanding of the circumstances of patient non-adherence.

Future work would include the investigation of alternative methods for modeling longitudinal data. Furthermore, alternatives to ratios such as PDC should be considered, as they may not provide the most accurate portrayal of medication adherence.

## BIBLIOGRAPHY

- [1] Elizabeth D Kantor, Colin D Rehm, Jennifer S Haas, Andrew T Chan, and Edward L Giovannucci. Trends in Prescription Drug Use Among Adults in the United States From 1999-2012. *JAMA*, 314(17):1818, November 2015.
- [2] Joyce A Cramer, Anuja Roy, Anita Burrell, Carol J Fairchild, Mahesh J Fuldeore, Daniel A Ollendorf, and Peter K Wong. Medication compliance and persistence: terminology and definitions. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 11(1):44–47, January 2008.
- [3] Niteesh K Choudhry, Soko Setoguchi, Raisa Levin, Wolfgang C Winkelmayr, and William H Shrank. Trends in adherence to secondary prevention medications in elderly post-myocardial infarction patients. *Pharmacoepidemiology and drug safety*, 17(12):1189–1196, December 2008.
- [4] Marsha A Raebel, Julie Schmittiel, Andrew J Karter, Jennifer L Konieczny, and John F Steiner. Standardizing Terminology and Definitions of Medication Adherence and Persistence in Research Employing Electronic Databases. *Medical Care*, 51:S11–S21, August 2013.
- [5] Elizabeth Whalley Buono, Bernard Vrijens, Hayden B Bosworth, Larry Z Liu, Leah L Zullig, and Bradi B Granger. Coming full circle in the measurement of medication adherence: opportunities and implications for health care. *Patient Preference and Adherence*, 11:1009–1017, June 2017.
- [6] Wei-Hsuan Lo-Ciganic, Julie M Donohue, Bobby L Jones, Subashan Perera, Joshua M Thorpe, Carolyn T Thorpe, Zachary A Marcum, and Walid F Gellad. Trajectories of Diabetes Medication Adherence and Hospitalization Risk: A Retrospec-

- tive Cohort Study in a Large State Medicaid Program. *Journal of general internal medicine*, 31(9):1052–1060, September 2016.
- [7] Alex J Adams. Defining and Measuring Primary Medication Nonadherence: Development of a Quality Measure. pages 1–8, April 2016.
- [8] Bradley C Martin, Elizabeth K Wiley-Exley, Shirley Richards, Marisa E Domino, Timothy S Carey, and Betsy Lynn Sleath. Contrasting Measures of Adherence with Simple Drug Use, Medication Switching, and Therapeutic Duplication. *Annals of Pharmacotherapy*, 43(1):36–44, January 2009.
- [9] David P Nau. Proportion of Days Covered (PDC) as a Preferred Method of Measuring Medication Adherence. *Pharmacy Quality Alliance*, pages 1–3, January 2012.
- [10] Andrew J Karter, Melissa M Parker, Howard H Moffet, Ameena T Ahmed, Julie A Schmittziel, and Joe V Selby. New Prescription Medication Gaps: A Comprehensive Measure of Adherence to New Prescriptions. *Health Services Research*, 44(5p1):1640–1661, October 2009.
- [11] Wai Yin Lam and Paula Fresco. Medication Adherence Measures: An Overview. *BioMed research international*, 2015:217047, 2015.
- [12] Jacqueline G Hugtenburg, Lonneke Timmers, Petra Jm Elders, Marcia Vervloet, and Liset van Dijk. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. *Patient Preference and Adherence*, 7:675–682, 2013.
- [13] Melissa M Parker, Howard H Moffet, Alyce Adams, and Andrew J Karter. An algorithm to identify medication nonpersistence using electronic pharmacy databases. *Journal of the American Medical Informatics Association : JAMIA*, 22(5):957–961, September 2015.

- [14] Alexandra Lelia Dima and Dan Dediu. Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data. *PloS one*, 12(4):e0174426, 2017.
- [15] Andrew M Peterson, David P Nau, Joyce A Cramer, Josh Benner, Femida Gwadry-Sridhar, and Michael Nichol. A checklist for medication compliance and persistence studies using retrospective databases. *Value in Health*, 10(1):3–12, January 2007.
- [16] Marie Hudson, Elham Rahme, Hugues Richard, and Louise Pilote. Comparison of measures of medication persistency using a prescription drug database. *American Heart Journal*, 153(1):59–65, January 2007.
- [17] William C Gerth. Compliance and persistence with newer antihypertensive agents. *Current hypertension reports*, 4(6):424–433, December 2002.
- [18] Daniel S Nagin, Bobby L Jones, Valéria Lima Passos, and Richard E Tremblay. Group-based multi-trajectory modeling. *Statistical Methods in Medical Research*, pages 096228021667308–9, December 2017.
- [19] William M Vollmer, Maochao Xu, Adrienne Feldstein, David Smith, Amy Waterbury, and Cynthia Rand. Comparison of pharmacy-based measures of medication adherence. *BMC health services research*, 12(1):155, June 2012.
- [20] J F Steiner and A V Prochazka. The assessment of refill compliance using pharmacy records: methods, validity, and applications. *Journal of clinical epidemiology*, 50(1):105–116, January 1997.
- [21] Dan Dediu, Alexandra Dima. *AdhereR: Adherence to Medications*, 2017. R package version 0.1.0.
- [22] Susan E Andrade, Kristijan H Kahler, Feride Frech, and K Arnold Chan. Methods for evaluation of medication adherence and persistence using automated databases.

- Pharmacoepidemiology and drug safety*, 15(8):565–74– discussion 575–7, August 2006.
- [23] Michael Taitel, Leonard Fensterheim, Heather Kirkham, Ryan Sekula, and Ian Duncan. Medication days' supply, adherence, wastage, and cost among chronic patients in Medicaid. *Medicare & medicaid research review*, 2(3), 2012.
- [24] Marsha A Raebel, Nikki M Carroll, Jennifer L Ellis, Emily B Schroeder, and Elizabeth A Bayliss. Importance of including early nonadherence in estimations of medication adherence. *Annals of Pharmacotherapy*, 45(9):1053–1060, September 2011.
- [25] Isabelle Arnet, Ivo Abraham, Markus Messerli, and Kurt E Hersberger. A method for calculating adherence to polypharmacy from dispensing data records. *International journal of clinical pharmacy*, 36(1):192–201, February 2014.
- [26] R A Greevy, M M Huizinga, C L Roumie, C G Grijalva, H Murff, X Liu, and M R Griffin. Comparisons of persistence and durability among three oral antidiabetic therapies using electronic prescription-fill data: the impact of adherence requirements and stockpiling. *Clinical pharmacology and therapeutics*, 90(6):813–819, December 2011.
- [27] Keith Dunn, Marie-Hélène Lafeuille, Xiaolong Jiao, Hela Romdhani, Bruno Emond, Kimberly Woodruff, Jacqueline Pesa, Neeta Tandon, and Patrick Lefebvre. Risk Factors, Health Care Resource Utilization, and Costs Associated with Nonadherence to Antiretrovirals in Medicaid-Insured Patients with HIV. *Journal of managed care & specialty pharmacy*, pages 1–12, June 2018.
- [28] Parya Saberi, Mallory O Johnson, Charles E McCulloch, Eric Vittinghoff, and Torsten B Neilands. Medication adherence: tailoring the analysis to the data. *AIDS and behavior*, 15(7):1447–1453, October 2011.



- [29] Richard A Kronmal. Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 156(3):379, 1993.
- [30] Jason Yeaw, Joshua S Benner, John G Walt, Sergey Sian, and Daniel B Smith. Comparing adherence and persistence across 6 chronic medication classes. *Journal of managed care pharmacy : JMCP*, 15(9):728–740, November 2009.
- [31] Sudeep Karve, Mario A Cleves, Mark Helm, Teresa J Hudson, Donna S West, and Bradley C Martin. Good and poor adherence: optimal cut-point for adherence measures using administrative claims data. *Current medical research and opinion*, 25(9):2303–2310, September 2009.
- [32] Bimal V Patel, R Scott Leslie, Patrick Thiebaud, Michael B Nichol, Simon S K Tang, Henry Solomon, Dennis Honda, and JoAnne M Foody. Adherence with single-pill amlodipine/atorvastatin vs a two-pill regimen. *Vascular health and risk management*, 4(3):673–681, 2008.
- [33] Joshua S Benner, Richard H Chapman, Allison A Petrilla, Simon S K Tang, Noah Rosenberg, and J Sanford Schwartz. Association between prescription burden and medication adherence in patients initiating antihypertensive and lipid-lowering therapy. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, 66(16):1471–1477, August 2009.
- [34] Frank E. Harrell. *Regression Modeling Strategies*. Springer, 2015.
- [35] Clay Ford. *Getting started with Negative Binomial Regression Modeling*, 2016. <http://data.library.virginia.edu/getting-started-with-negative-binomial-regression-modeling/> [Accessed: 2018-05-11].

- [36] NCSS Statistical Software. *Chapter 326: Negative Binomial Regression*. [https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative\\_Binomial\\_Regression.pdf](https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Negative_Binomial_Regression.pdf) [Accessed: 2018-05-11].
- [37] *Negative Binomial Regression: R Data Analysis Examples*, 2013. <https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/> [Accessed: 2018-05-11].
- [38] Jonathan Schildcrout. Applied Regression Analysis II, Module 3, Marginal Models and GEE. BIOS 7346 Lecture Notes, 2018.