Modeling Zero-Inflated and Overdispersed Count Data With Application to Psychiatric Inpatient Service Use

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the School of Public Health
University of Saskatchewan
Saskatoon

By
Sharmin Sharker

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Director of School of Public Health
Health Sciences Building E-Wing, 104 Clinic Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 2Z4
Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9

Canada

Abstract

Psychiatric disorders can be characterized as behavioral or mental states that cause significant distress and impaired personal functioning. Such disorders may occur as a single episode or persistent, relapsing, and perhaps leading to suicidal behaviours. The exact causes of psychiatric disorders are hard to determine but easy access to health care services can help to reduce the severity of the states. Inpatient psychiatric hospitalization is not only an expensive mode of treatment but also may represent the quality of health care system. The aim of this study was to investigate the factors associated with repeated hospitalizations among the patients with psychiatric illness, which may help the policy makers to target the high-risk groups in a more focused manner.

The count of hospitalizations for psychiatric patients may be zero during a period of time for the huge majority of patients rather than a positive count. A common strategy to handle excessive zeros is to use zero-inflated models or hurdle models. In the field of health services research of mental health, very little literature is available comparing the relative fits of zero-inflated distributions and other count distributions to empirical data. A large linked administrative database consisting of 200,537 patients with psychiatric diagnosis in the years of 2008-2012 was used in this thesis. Various counts regression models were considered for analyzing the hospitalization rate among patients with psychiatric disorders within 3, 6 and 9 months follow-up since index visit date. The covariates for this study consist of socio-demographic and clinical characteristics of the patients. According to the Akaike Information Criteria, Vuong's test and randomized quantile residuals, the hurdle negative binomial model was the best model.

Our results showed that hospitalization rate depends on the patients' socio-demographic characteristics and also on disease types. It also showed that having previously visited a general physician served a protective role for psychiatric hospitalization during our study period. Patients who had seen an outpatient psychiatrist were more likely to have a higher number of psychiatric hospitalizations. This may indicate that psychiatrists tend to see patients with more severe illnesses, who require hospital-based care for managing their illness. Having earlier and greater access to outpatient psychiatrist and community-based mental

health care may alleviate the need for hospital-based psychiatric care.

Acknowledgements

I would like to first and foremost express my sincere gratitude to my supervisors, Dr. Cindy Xin Feng and Dr. Lloyd Balbuena, for their academic and financial support as well as their encouragements and patience throughout the period of my study. Without their support and motivations, enlightening advice, impressive kindness and patience, I could not have completed my thesis. It is a great honour to work under their supervision.

I would like to thank my thesis advisory committee members: Dr. Punam Pahwa, Dr. Bonnie Janzen and Dr. Shahedul Khan and my external examiner: Dr. Mobinul Huq for their very insightful suggestions and comments of my thesis. I also acknowledge the Collaborative Biostatistics Program at the School of Public Health of University of Saskatchewan for academic and financial support to my MSc study.

Finally, I would like to thank Dr. Gene Marcoux for allowing me to use the data acquired from the Ministry of Health.

I want to state that this Study is based in part on de-identified data provided by the Saskatchewan Ministry of Health. The interpretation and conclusions contained herein do not necessarily represent those of the Government of Saskatchewan or the Saskatchewan Ministry of Health.

Contents

List of Tables

List of Figures

1. Introduction

1.1 Background and Rationale

Mental illness causes suffering for as many as 450 million people worldwide at any given period of their life time [1]. One in five Canadians will live with a mental illness in their lifetimes [2]. The increased demand for health care for mental health concerns has been identified as an important public health topic all over the world and in Canada as well. Timely and easy access to care are an essential aspects of mental health care quality. However, it is possible that most mental health patients can not receive timely outpatient care which may contribute to increased demand for acute hospital care. The aim of our study was to make a significant contribution in the context of health care utilization research in mental health, since few research studies have been conducted to study the factors associated with inpatient mental health hospitalizations in Saskatchewan.

Inpatient mental health hospitalizations represent a fraction of the spectrum of services in place for those who seek treatment for mental illness [3]. Inpatient hospitalization is often resorted to when a mental condition worsens and treatment in an outpatient setting is insufficient. In health services research, inpatient psychiatric hospitalizations are an expensive mode of treatment and readmission within a short period of time is a negative performance indicator [3–5] and a major driver of cost. Densen et al. [6] call individuals who account for a large portion (50-70%) of health spending high-cost users and Taube et al. [7] demonstrated that individuals with mental disorders are responsible for a disproportionate amount of outpatient expenses. Several recent Canadian studies have focused on high-cost users and patients with mental health and addiction (MHA) issues. De Oliveira et al. [8] demonstrated that high-cost MHA patients incur 30% more healthcare costs per capita compared to high-cost users with no mental health conditions. Likewise, a subsequent study demonstrated that MHA high-cost patients (i.e. individuals whose MHA services accounted for more than 50% of their total healthcare costs) had healthcare costs 40% higher than those with no MHA-

related costs [9]. Adopting a combination of mood, substance use, psychotic and anxiety disorders as a definition of mental illness, Hensel et al. [10] found that the prevalence of mental illness was 39.3% in the top 1% of users by cost (compared to 21.3% in the lowest cost group). Among MHA high-cost users, people with schizophrenia are common, largely driven by their need for frequent hospitalizations [10–12].

Aside from being costly, frequent inpatient mental health care may also reflect the underlying deficiencies in the availability or access to outpatient care services for pre- and post-discharge treatment [13] and further a negative indicator of the quality of hospital care [14, 15]. Several strategies have been developed for reducing readmission rates [16]. These include enhanced patient education, more intensive post-discharge follow-up care, and increased coordination with outpatient providers [17]. A study based in Saskatchewan reported that having a good connection to a primary care provider decreased the probability of being a high-cost health service user [18]. Nevertheless, lack of access to outpatient mental health care may not be the only factor contributing to increased demand for inpatient care. Elucidating factors that contribute to inpatient mental health hospitalizations is essential for further understanding the population heterogeneity of mental health seeking behaviors. Identification of population subgroups with frequent hospitalization for mental health concerns may help public health policymakers to target those sub- populations in a more focused and efficient manner.

Previous research indicated that inpatient hospitalizations for patients with psychiatric disorders may reflect both the severity and type of illness [19]. Inpatient readmission is common for individuals with severe mental illness (SMI) (e.g., schizophrenia, mood disorders, bipolar disorder and psychoses) with estimates in the range of 40% to 54% [20–23]. It was found in England that the life expectancy of SMI patients is 10 to 15 years shorter than the general population [24]. A recent global morbidity study attributed 3.5 % of total Years Lost to Disability to two types of diseases: schizophrenia and bipolar disorder combined [25]. These two diseases alone are estimated to constitute 1.1% of the total Disability Adjusted Life Year burden of disease in 21 regions worldwide [26]. People with SMI are at higher risk of hospitalizations compared to the general population [27, 28] partly because medical comorbidities are more common [29, 30]. SMI is associated with increased treatment costs and

hospitalizations may represent a significant proportion of their health care costs. Predicting inpatient hospitalization is difficult due to the complex interplay between systemic, clinical, and sociodemographic factors. Many studies have focused on patient risk factors for inpatient hospitalization, including severity of psychotic symptoms, substance use, suicidal behavior, and current manic symptoms [31–35].

In addition to clinical factors, demographic and socio-economic factors have been shown to be associated with mental health services utilization. Recent studies indicated that mental health problems and mental health service utilization among children and youth have been increasing over the last three decades [36, 37]. A survey conducted in Ontario by Offord et al. [38] showed that about one in five (18.1%) children between the ages of 4 and 16 years experience at least one of the following psychiatric disorders (conduct disorder, hyperactivity, emotional disorder, and somatization). The prevalence rates of mental disorders among children and youth in Canada was estimated to be about 14% [39]. Mason and Gibbs [40] examined hospitalization patterns among adolescents and found that younger children were significantly more likely to experience a longer hospitalization than the older ones. Fontanella [41] conducted a 1-year longitudinal study on children and youth discharged from psychiatric hospital care and found that the lack of parental involvement and use of corporal punishment predicted psychiatric readmission. Also, in a 2.5-year longitudinal study by James et al. [42], the researchers found that availability of post-discharge services reduced the risk of rehospitalization among children and youth. Other research (e.g. [43]) found that American children and youth in state custody who were living in group homes or youth emergency shelters were more likely to have been readmitted. However, studies of elderly populations have reported equivocal findings regarding the associations between depressive symptoms and hospitalization episodes, with some reporting a lack of association [44–46], and others reporting a positive association, but only among subgroups of elderly men [47]. CIHI [48] reported that older individuals are more likely to be rehospitalized. This study was based on the patterns of one-year readmissions to acute care hospitals in Canada among patients who had a mental illness as the most responsible diagnosis in their index admission during 2002-2003.

Carriere et al. [49] found that hospitalization rates for mental illness were almost in-

variably higher for First Nations living on and off reserve, Metis, and Inuit than for the non-Aboriginal population, regardless of disease category. One possible explanation of this situation provided by Carrier and Pong [49, 50], was that on-reserve First Nations people primarily live in rural areas and may have less access to primary care services. Females were found to be higher users of high-cost health facility users for psychiatric illness more than males [18]. WHO [51] claimed that depression is two times more common among females than males.

Residential locations may also play an important role in hospitalization with geographical characteristics such as population and service location being considered as important factors in several studies. The most commonly used procedures included area level variables like rural or urban areas, or the availability of community care in that area as predictors or compared different geographical areas [14, 15]. Nevertheless, the research findings are mixed. Some studies reported that readmission rates were lower in urban regions [52, 53]; whereas, positive associations between readmission rates and population density were reported in other studies [52].

1.2 Review of Methodology

Hospitalization rates (i.e. number of hospitalizations per unit time) are a measure of disease burden and a key driver of health care costs. Many studies have focused on the population with at least one hospitalization admission without considering the relatively healthy subpopulation who had no hospitalizations within a certain study period. A comparison between readmitted versus not readmitted patients was typically performed using logistic regression. Priebe [54] considered the readmission rate per person-year, while in other cases separate analyses were performed for psychiatric and non-psychiatric reasons [55]. Various comparisons involved patients readmitted during a given time period vs a control group of non-readmitted patients [56], early vs late readmission vs control patients [57, 58], or readmitted vs several groups of non-readmitted [59] (community and nursing home). Several studies have simply compared those patients who have been readmitted versus not readmitted–ignoring number of readmissions. The resulting two groups were then compared using basic statistical tech-

niques such as the chi-squared test [60], or by more sophisticated techniques such as logistic regression [61, 62]. However, the use of such dichotomization leads to pertinent information being ignored, such as the distribution of readmissions. To address this weakness, other studies [63, 64] resorted to using linear regression to model the number of readmissions. It is well-known that applying linear regression to count data, such as number of readmissions, is problematic[65]. In particular, the significance tests and confidence intervals employed in linear regression assume equal variance, which are typically violated by count data.

Poisson regression is commonly used for analyzing counts data [66–69]. However, the Poisson distribution assumes that the mean and variance are equal conditional on a given set of covariate values, [70–72]. Poisson regression may not perform well when the conditional variance is greater than the conditional mean [73]. Such a phenomenon is known as overdispersion. Eaton [74] analysed 2316 hospital admissions for schizophrenia, based upon all the psychiatric hospitals in the state of Maryland. It was found that the readmission data were fitted poorly by the Poisson distribution. Similar results were found by Smeeton [75] who analysed 59,546 episodes of psychiatric illness recorded by general practitioners in England and Wales. In both cases, the conditional variance for the number of readmissions or episodes was higher than the conditional mean, indicating the presence of overdispersion. The negative binomial (NB) distribution could serve as the basis for modeling data with overdispersion [65]. The NB distribution was originally applied to the study of multiple accidents or illnesses [76] but has also been applied to the study of health service and resource utilization including visits to clinics [77], intervals between suicides [78] and acts of violence by psychiatric outpatients [65]. Eaton [74] and Smeeton [75] both found that the NB distribution fit the data well. Moreover, in a general hospital study, a study analysed 458,563 admissions of New England residents aged 65 or older [79] and found that the NB regression model fit the data far better than did the Poisson model. Using the latter distribution led to results that were too liberal or optimistic, in that the confidence intervals [80] for the regression parameters were 25% to 50% smaller than those obtained using the NB distribution. In other words, the inappropriate use of Poisson regression could lead to unjustifiably precise (narrow) confidence intervals for the regression parameters.

In situations when a large number of patients are not hospitalized during the study

period, neither Poisson or NB regressions may fit the data well [81]. Alternative modeling techniques have been proposed in literature to accommodate the excess zeros, including hurdle and zero-inflated models. A hurdle model [82] is a two-component model in which one component models the probability of zero counts and the other component uses a zero-truncated Poisson or zero-truncated NB distribution that is modified by conditioning on a positive outcome. Another way to deal with excessive zeros is to use a zero-inflated model [83], which is a mixture of a regular count regression model such as Poisson or NB model and a component that accommodates the excessive zeros. The difference between hurdle and zero-inflated models is that a hurdle model considers the zeros to be separate from the non-zeros, whereas for a zero-inflated model, there are two distinct processes by which the zeros arise: sampling zeros which occur by chance and structural zeros or true zeros which are inevitable and are part of the counting process.

Modeling hospitalization rate requires defining a study period. The literature on this subject is difficult to interpret as it is composed of studies examining readmission within periods ranging from 1 month to 1 year conducted in a wide variety of settings with different patient populations, different models of inpatient care and community care, and these differences may have led to contradictory conclusions [21]. In a study by Lyons et al. [84], the success of a hospital intervention, assessed by the improvement in psychiatric symptoms, had no influence on readmission in the 30 days or 6 months following discharge. This was true for a wide range of psychiatric disorders. Lyons et al.[84] focused on short-term readmission (i.e. 30 days and 6 months) which is based on the premise that a particular hospitalization decreases its impact on readmission relative to other variables such as environmental stressors [85]. We will therefore conduct a sensitivity analysis by varying the follow-up time from 3, 6, to 9 months to test the robustness of our findings.

1.3 Study Motivation

For modeling counts data, the choice of underlying distribution is crucial for valid statistical inference, which commonly depends on the percentage of zeros, the underlying data generating processes by which zeros arise and dispersion of the positive data. In the field of

mental health services research of mental health, very little literature is available comparing the relative fits of zero-inflated distributions and other count distributions to empirical data.

In this thesis, two important questions were addressed, relating to the choice of statistical models: Do Poisson, NB, zero-inflated and hurdle models yield different model fits for modeling repeated hospitalization rate? And, what factors are significantly associated with hospitalization rate?

Model fits are compared in terms of:

- Relative fits measures: Akaike's information criterion (AIC) [86]. Vuong's test [87], a likelihood-ratio-based test for model selection.
- Absolute fits measures: Random quantile residuals (RQR) [88], and, the fitted values vs observed values.

Our study aimed to identify the significant socio-demographic and clinical characteristics for the individuals associated with the inpatient hospitalizations. In particular, our objective was to assess whether access to primary care and psychiatric outpatient services protects against mental health hospitalization. Mental health services outside of the hospital may reduce the need for hospitalizations, so this study would provide valuable information for policymakers and planners. For example, planners would be able to identify health regions with higher hospitalization rates and take proper steps to address modifiable risks for hospitalization.

1.4 Organization of the Thesis

Different modeling methods for counts data are reviewed in Chapter 2. In Chapter 3, model comparison and diagnosis methods for counts regression models are discussed. Chapter 4 presents the analysis of the inpatient hospital care use for psychiatric patients data using various counts regression models. Model fits are compared using various model comparison and goodness of fit criteria. Finally, Chapter 5 presents the concluding remarks of this thesis.

2. Methodology

In this Chapter, we briefly review some commonly used regression models for count data, including Poisson and negative binomial (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), hurdle Poisson (HP) and hurdle negative binomial (HNB) models.

2.1 Statistical Models for Modeling Counts Data

2.1.1 Poisson Model

Poisson distribution is often employed for analyzing count data [66–69]. One of the important assumptions of the Poisson distribution is that the mean and variance have to be equal, which is very restrictive [70–72]. Overdispersion occurs if the variance is greater than the mean [89].

Let $Y_i, i = 1, ..., n$ be a discrete random variable which follows a Poisson distribution with parameter μ_i, and n be the sample size. The probability mass function (PMF) for a Poisson distribution is

$$f(Y_i = y_i; \mu_i) = \frac{(e^{-\mu_i})(\mu_i^{y_i})}{y_i!}, y_i = 0, 1, ..., n. \tag{2.1}$$

The expected mean and variance of Y_i are as follows:

$$E(Y_i) = \mu_i \tag{2.2}$$

$$V(Y_i) = \mu_i. \tag{2.3}$$

Typically, a logarithm is used to link μ_i to a linear predictor of X_i

$$\log(\mu_i) = X_i^T \beta, \tag{2.4}$$

where X_i represents a vector of explanatory variables for the ith subject $(1, X_{i1}, X_{i2}, ..., X_{ip})^T$ and $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$ is a vector of regression coefficients.

2.1.2 Negative Binomial Model

Negative binomial (NB) distribution is an extension of the Poisson distribution, where the Poisson mean μ_i is distributed as a gamma distribution. Detailed discussions can be found in [90, 91].

Suppose Y_i has a NB distribution with mean μ_i and shape parameter k. The shape parameter k controls overdispersion.

The probability mass function can be written as,

$$f(Y_i = y_i; \mu_i, k) = \frac{\Gamma(\mu_i + k)}{\Gamma(k)\Gamma(\mu_i)}(\frac{\mu_i}{\mu_i + k})^{y_i}(\frac{k}{\mu_i + k})^k, y_i = 0, 1, ..., n. \tag{2.5}$$

The mean and variance of y_i are

$$E(Y_i) = \mu_i \tag{2.6}$$

$$V(Y_i) = \mu_i + \frac{\mu_i^2}{k}. \tag{2.7}$$

If k tends to infinity, the NB is reduced to the Poisson distribution. If $k > 0$ then $V(Y_i) > E(Y_i)$. Typically a logarithm is used to link μ_i to a linear predictor of X_i

$$\log(\mu_i) = X_i^T \beta, \tag{2.8}$$

where X_i represents a vector of explanatory variables for the ith subject $(1, X_{i1}, X_{i2}, ..., X_{ip})^T$ and $\beta =(\beta_0, \beta_1, \beta_2, ..., \beta_p)^T$ is a vector of regression coefficients.

2.1.3 Zero-Inflated Model

The zero-inflated model was firstly proposed by Lambert [92] with an application to defects in a manufacturing process. One assumption of this model is that structural zeros occur with probability p_i, and sampling zeros with probability $(1 - p_i)$. Thus, the occurrence of y_i from a zero-inflated Poisson (ZIP) model follows the following distribution:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i} \text{ if } & y_i = 0 \\ (1 - p_i)\frac{(e^{-\mu_i})(\mu_i^{y_i})}{y_i!} \text{ if } & y_i > 0 \end{cases}, \tag{2.9}$$

The mean and variance of ZIP distribution are

$$E(Y_i) = (1 - p_i)\mu_i \tag{2.10}$$

$$V(Y_i) = (1 - p_i)(\mu_i + p_i\mu_i^2). \tag{2.11}$$

This distribution approaches to Poisson(μ_i) as $p_i \rightarrow\ 0$.

The ZIP model with log link function is:

$$\begin{cases} \log(\mu_i) = X_i^T\beta \\ \text{logit}(p_i) = \log(\frac{p_i}{1-p_i}) = \tilde{X}_i^T\tilde{\beta}, \end{cases} \tag{2.12}$$

where $0 \leq p_i \leq 1$ and $0 < \mu_i < \infty$. $\tilde{X}_i$ represents a set of covariates related to the binary outcome and $\tilde{\beta}$ represents the respective estimates.

Zero-inflated negative binomial (ZINB) distribution is a mixture distribution, similar to ZIP distribution, where p_i denotes the probability that an individual belongs to the excess zero component and with probability $(1-p_i)$, the rest of the counts follow a NB distribution.

The ZINB distribution is given by:

$$P(Y_i = y_i; \mu_i, k) = \begin{cases} p_i + (1 - p_i)(1 + \frac{\mu_i}{k})^{\ -k}\ ,\ y_i = 0 \\ (1 - p_i)\frac{\Gamma(y_i+k)}{\Gamma(k)\Gamma(y_i)}(\frac{\mu_i}{\mu_i+k})^{y_i}(\frac{k}{\mu_i+k})^k\ ,\ y_i = 1, 2, ... \end{cases} \tag{2.13}$$

The mean and variance of ZINB distribution are

$$E(Y_i) = (1 - p_i)\mu_i \tag{2.14}$$

$$V(Y_i) = (1 - p_i)\mu_i(1 - p_i\mu_i + \frac{\mu_i}{k}). \tag{2.15}$$

Note that this distribution approaches the ZIP distribution and the NB distribution as $k \rightarrow \infty$ and $p_i \rightarrow 0$, respectively. If both $\frac{1}{k}$ and $p_i \approx 0$ then ZINB distribution reduces to Poisson distribution. The ZINB mixed effects model can be written similarly as equation (2.12) for a ZIP model.

2.1.4 Hurdle Model

In a hurdle model, the first part is a binary response model and the second part is usually a truncated-at-zero count model [93]. Hence, a hurdle model is a modified count model in

which separate processes generating the zeroes and positive counts are not constrained to be the same. This allows us to interpret the positive outcomes (>0) that result from passing the zero hurdle (threshold). The hurdle portion of the two-part model estimates the probability that the threshold is crossed. Theoretically, the threshold could be any value, but it is usually set equal to zero because this is most often meaningful in the context of the study objectives. Mullahy [82] laid out the basic form of hurdle count models.

The unconditional PMF for Y is given by

$$Pr(Y_j = y_i) = \begin{cases} \pi_i & \text{if } y_i = 0 \\ (1-\pi_i)\frac{f_1(y_i)}{1-\ f_1(0)} & \text{if } y_i > 0. \end{cases} \tag{2.16}$$

where π_i is the probability of zeroes. The mean and variance for the hurdle Poisson (HP) model are:

$$E(Y_i) = \frac{1-\pi_i}{1-e^{-\mu_i}} \times \mu_i \tag{2.17}$$

$$V(Y_i) = \frac{1-\pi_i}{1-e^{-\mu_i}} \times \left(\mu_i + \mu_i^2\right) - \left(\frac{1-\pi_i}{1-e^{-\mu_i}} \times \mu_i\right)^2. \tag{2.18}$$

The HP model with the log link function for the truncated Poisson component and the binomial component with the logit link function are then written as:

$$\begin{cases} \log(\mu_i) = X_i^T\beta \\ \text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \tilde{X}_i^T\tilde{\beta}, \end{cases} \tag{2.19}$$

HNB model can be defined analogously [94].

2.1.5 Counts Regressions for Modeling Event Rate

For longitudinal follow-up data, study participants may be subject to loss to follow-up. Therefore, it is important to model the number of repeated hospitalizations accounting for varying follow-up times. For example, the number of repeated hospitalizations could be much higher for those participants who had 9 months of follow up as compared to those who had 3 months of follow up due to the longer observational period. In this study, we use the follow up time (τ_i) as the offset term which is random. The response variable (i.e., the number of

repeated hospitalizations) was divided by the offset variable as the adjustment, which is the hospitalization rate.

For a Poisson or NB model, we then model the "event rate", which is the expected number of events per unit time as,

$$\theta_i = \frac{\mu_i}{\tau_i}, \tag{2.20}$$

Therefore, a Poisson or NB model can be written as:

$$\log(\mu_i) = \log(\tau_i) + \log(\theta_i) = \log(\tau_i) + X_i^T \beta. \tag{2.21}$$

A ZI model for modeling the "event rate" is then written as:

$$\begin{cases} \log(\mu_i) = \log(\tau_i) + X_i^T \beta \\ \text{logit}(p_i) = \log(\tau_i) + \tilde{X}_i^T \tilde{\beta}. \end{cases} \tag{2.22}$$

Similarly, a hurdle model for modeling the "event rate" can be formulated as:

$$\begin{cases} \log(\mu_i) = \log(\tau_i) + X_i^T \beta \\ \text{logit}(\pi_i) = \log(\tau_i) + \tilde{X}_i^T \tilde{\beta}, \end{cases} \tag{2.23}$$

where $\log(\tau_i)$ is the "offset" term to account for varying exposure/follow-up times among the study participants.

2.2 Estimation Procedure

There are a few methods for estimating the parameters of a generalized linear model (GLM). Maximum likelihood estimation technique is easily applicable to estimate the parameters of a GLM. Moreover, maximum likelihood estimates can be used to estimate the variance components from the second derivative of the log likelihood. Newton-Raphson, Nelder-Maid and similar algorithms are available to maximize the complete data loglikelihood function. Package `glmmTMB` can be used for fitting generalized linear models and extensions. This is an `R` package built on the Template Model Builder automatic differentiation engine. The distributions for the response variable that are considered in `glmmTMB` include: Poisson, binomial, negative binomial, Gamma, Beta, Gaussian, zero-truncated Poisson and zero-truncated negative binomial. This package also allows to add offset in the models.

The general likelihood function of HP model can be written as:

$$\begin{aligned} L(\beta,\tilde{\beta}) &= \prod_{(y_i=0)}\left[P(y_i=0)\right]\prod_{(y_i>0)}\left[P(y_i>0)f_1(y_i)\right] \\ &= \left[\prod_{(y_i=0)}\tfrac{1}{1+e^{x_i^T\beta}}\right]\left[\prod_{(y_i>0)}\tfrac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\tfrac{e^{-e^{\tilde{x}_i^T\tilde{\beta}}}e^{y_i\tilde{x}_i^T\tilde{\beta}}}{(1-e^{-e^{\tilde{x}_i^T\tilde{\beta}}})y_i!}\right] \\ &= \left[\prod_{(y_i>0)}e^{x_i^T\beta}\prod_{i=1}^{n}\tfrac{1}{1+e^{x_i^T\beta}}\right]\left[\prod_{(y_i>0)}\tfrac{e^{-e^{\tilde{x}_i^T\tilde{\beta}}}e^{y_i\tilde{x}_i^T\tilde{\beta}}}{(1-e^{-e^{\tilde{x}_i^T\tilde{\beta}}})y_i!}\right] \end{aligned} \tag{2.24}$$

The likelihood function for HNB can also be written in the same way as HP.

The likelihood function for ZIP can be written as:

$$\begin{aligned} L(\beta,\tilde{\beta}) &= \prod I_{(y_i=0)}\left[p_i+(1-p_i)e^{-\mu_i}\right]I_{(y_i>0)}\left[(1-p_i)\frac{(e^{-\mu_i})(\mu_i^{y_i})}{y_i!}\right] \\ &= \prod\left[\left(\frac{e^{\tilde{x}_i^T\tilde{\beta}}}{1+e^{\tilde{x}_i^T\tilde{\beta}}}+\frac{e^{-e^{x_i^T\beta}}}{1+e^{\tilde{x}_i^T\tilde{\beta}}}\right)+\frac{e^{-e^{x_i^T\beta}}(e^{x_i^T\beta})^{y_i}}{(1+e^{\tilde{x}_i^T\tilde{\beta}})y_i!}\right] \end{aligned} \tag{2.25}$$

where, $I_{(.)}$ =1 if the condition is true, 0 otherwise. The likelihood function for ZINB can also be written in same manner as ZIP.

2.3 Hypothesis Testing

The very first step in conducting a study is to specify a proper research hypothesis. A null hypothesis (H_0), which is a theoretical statement that no relationship exists between the exposure or intervention and the outcome of interest, i.e. it is the hypothesis of no difference. If there is a significant departure in the data from what would be expected under H_0 , the investigator is able to reject the hypothesis in favour of an alternative hypothesis (H_A), that an association exists between the exposure or intervention and the outcome.

There are three standard ways to use the likelihood function to perform large-sample inference [95].

Wald Test

Let us consider β be an arbitrary parameter and $H_0 : \beta = \beta_0$. If the $\hat{\beta}$ is the unrestricted Maximum likelihood (ML) estimates of β, then we can write the test statistics under the H_0

as [95]:

$$z = (\hat{\beta} - \beta_0)/SE(\hat{\beta}), \tag{2.26}$$

where $SE(\hat{\beta})$ denotes the standard error of $\hat{\beta}$.

This statistic has an approximate standard normal distribution under the H_0. z is referred to the standard normal table to obtain one or two-sided P-values [95]. Equivalently, for the two-sided alternatives, z^2 has an approximate chi-squared null distribution with df=1; the P-value is then the right-tailed chi-squared probability above the observed value. This type of statistic, is called a *Wald statistic* [96].

Likelihood Ratio Test

A second general-purpose method uses the likelihood function through the ratio of two maximizations: (1) the maximum over the possible parameter values under H_0, and (2) the maximum over the larger set of parameter values permitting H_0 or an alternative H_a to be true [95]. Let l_0 denote the maximized value of the likelihood function under H_0, and let l_1 denote maximized value generally. For instance, for parameters $\beta = (\beta_0, \beta_1)$ and $H_0 : \beta_0 = 0$, l_1 is the likelihood function calculated at the β value for which the data would have been most likely; l_0 is the likelihood function calculated at the β_1 value for which the data would have been most likely, when $\beta_0 = 0$. Wilks showed that -2log(l_0/l_1) has a limiting null chi-squared distribution, as $n \to \infty$ [95]. The *likelihood-ratio test statistic* equals:

$$-2log(l_0/l_1) = -2(L_0 - L_1), \tag{2.27}$$

where L_0 and L_1 denote the maximized log-likelihood functions.

Score Test

The third method uses the *score statistic*. The score test is based on the slope and expected curvature of the log-likelihood function $L(\beta)$ at the null value β_0 [95]. It utilizes the size of the *score function*

$$U(\beta) = \partial L(\beta)/\partial(\beta). \tag{2.28}$$

Let $I(\beta)$ denote the Fisher information $-E[\partial^2 L(\beta)/\partial\beta^2]$ [95]. The chi-squared form of the score statistic is defined as:

$$\frac{[U(\beta_0)]^2}{I(\beta_0)} = \frac{[\partial L(\beta_0)/\partial\beta_0]^2}{-E[\partial^2 L(\beta_0)/\partial\beta_0^2]}, \tag{2.29}$$

which has an asymptotic distribution of χ_1^2 when $\beta = \beta_0$.

Confidence Interval

For any of the three tests mentioned earlier, a confidence interval can be constructed by inverting the test. For instance, a 95% confidence interval for β is the set of β_0 for which the test of $H_0 : \beta = \beta_0$ has P-value exceeding 0.05 [95].

Let z_α denote the z-score from the standard normal distribution having right-tailed probability α; this is the 100(1-α) percentile of that distribution [95]. A 100(1-α)% confidence interval based on asymptotic normality uses $z_{\alpha/2}$. The Wald confidence interval is the set of β_0 for which $\left|\hat{\beta} - \beta_0\right| / SE(\hat{\beta}) < z_{\alpha/2}$. This gives the interval $\hat{\beta} \pm z_{\alpha/2} SE(\hat{\beta})$ [95].

3. Model Selection and Goodness-of-Fit

In this Chapter, we introduce model selection methods and goodness-of-fit tests for comparing and diagnosing count regression models discussed in Chapter 2.

3.1 Model Selection

3.1.1 Information Criteria

Akaike information criterion (AIC) was first proposed by Akaike [86], which is used for comparing nested or non-nested models. AIC is given by:

$$\text{AIC} = -2\log L(\hat{\theta}) + 2c, \tag{3.1}$$

where $L(\hat{\theta})$ is the maximized likelihood function of a candidate model given the data when evaluated at the maximum likelihood estimate of θ and $-logL(\hat{\theta})$ offers summary information on how much discrepancy exists between the candidate model and the data, where c is the number of estimated parameters in the candidate model. AIC indicates goodness of fit and penalizes the number of parameters at the same time to avoid overfitting. The model with the lowest AIC value is preferred compared to other models.

A few modifications to AIC have been proposed. They impose different penalties for the number of parameters.

3.1.2 Vuong's Test

Vuong's test [87] is a likelihood-ratio-based test for model comparison in which the null hypothesis sets the two models equal to one another. The test statistic is given by

$$V = \frac{\bar{m}\sqrt{n}}{S_m} \tag{3.2}$$

with

$$m_i = \log\left[\frac{\hat{P}_1(Y_i|X_i)}{\hat{P}_2(Y_i|X_i)}\right], \tag{3.3}$$

where m_i is the log-likehood ratio between two models with $\hat{P}_1(Y_i|X_i)$ and $\hat{P}_2(Y_i|X_i)$ denoting the likelihood of two models. The statistic m_i has a mean $\bar{m}$ and standard deviation S_m and n is the sample size. The statistic V asymptotically follows a standard normal distribution. V greater than 1.96 supports $\hat{P}_1(Y_i|X_i)$ and V less than -1.96 supports the $\hat{P}_2(Y_i|X_i)$ at 5% level of significance.

3.2 Model Diagnosis

3.2.1 Randomized Quantile Residual (RQR)

To overcome the difficulties of using traditional residuals for diagnosing regression models for discrete outcomes, randomized quantile residual (RQR) [88] was proposed by inverting the fitted distribution function for each response value and finding the equivalent standard normal quantile. Let F(y_i;μ_i,ϕ) denote the cumulative distribution function (CDF) for random variable y. If the CDF is continuous, F(y_i; μ_i,ϕ) is uniformly distributed on the unit interval. RQRs can then be defined as

$$q_i = \Phi^{-1}[\ F(y_i; \hat{\mu}_i, \hat{\phi}_i)]\ , \tag{3.4}$$

where $\Phi^{-1}()$ is the quantile function of a standard normal distribution. However, if the CDF is discrete, randomization is added to make it continuous. To be more specific, let $p(y_i; \mu_i, \phi)$ denote the PMF of y_i. The CDF can be redefined as:

$$F^*(y_i; \mu_i, \phi, u_i) = \begin{cases} F(y_i; \mu_i, \phi), & F \text{ is continuous} \\ F(y_i^-; \mu_i, \phi) + u_i\, p(y_i; \mu_i, \phi), & F \text{ is discrete} \end{cases} \tag{3.5}$$

where u_i is a uniform random variable on $[0, 1]$, and $F(Y_i^-; \mu_i, \phi)$ is the lower limit of F in y_i. When F is discrete, we let $a_i = \lim_{y \to y_i^-} F(y; \hat{\mu}_i, \hat{\phi}_i)$ and $b_i = F(y_i; \hat{\mu}_i, \hat{\phi}_i)$, then the randomized quantile residual is

$$q_i = \Phi^{-1}(\ F_i^*), \tag{3.6}$$

where F_i^* is a uniform random variable on the interval $(a_i, b_i]$, and $q_i \sim N(0, 1)$. Therefore, the only information that is required for calculating RQRs is the CDF of the response variable.

3.2.2 Area Under the ROC Curve (AUC)

For a hurdle model, the logistic component can be diagnosed by the Receiver Operating Characteristics (ROC) curve, which is a standard technique for summarizing classifier performance over a range of trade-offs between true positive and false positive error rates [97]. A ROC curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR). For example, in the context of our motivating study, the true positive rate is the proportion of observations that were correctly predicted to be hospitalized out of all hospitalized patients (TP/(TP + FN)). False positive rate is the proportion of patients that are incorrectly predicted to be hospitalized out of all non-hospitalized patients (FP/(TN + FP)). The Area Under the Curve (AUC) measures the area under the ROC curve [98]. The maximum AUC=1 means that the model is perfect for distinguishing between the hospitalized and non-hospitalized patients. The minimum AUC should be considered a chance level, i.e. AUC=0.5. AUC can also represent the sensitivity and specificity of the analysis. Sensitivity is also known as the true positive rate, the recall, or probability of detection [99] in some fields. It measures the proportion of actual positives that are correctly identified as such (e.g., the probability of a patient being hospitalized can be correctly identified as being hospitalized). Specificity is also called the true negative rate. It measures the proportion of actual negatives that are correctly identified as such (e.g., the probability of a patient not being hospitalized can be correctly identified as not being hospitalized).

4. Data Analysis

4.1 Data Sources and Description

Data Source: The dataset consists of medical records for patients with psychiatric disorders from Jauary 1, 2008 to December 31, 2012. The sole criterion for inclusion into this cohort was a single medical services claim submitted to the Saskatchewan Medical Services Plan between January 1, 2010 and December 31, 2011 with an ICD-9 code from the mental disorders chapter (i.e., 290-319) reported in the diagnosis field. Past (January 1, 2008 to December 31, 2009) and future (2010-2012) hospital seperation records were then extracted for cohort members. The data was provided by Saskatchewan's Ministry of Health.

Person registry database: Each patient was assigned with a study ID number. The person registry information consists of the patients gender, year of birth, study entry date, study index date, study exit date, reason for exit, their registered Indian flag and residence at index date (defined below). The entry date is January 1, 2008 or coverage initiation with the Saskatchewan Ministry of Health. The index date is the date of the first physician service claim on the Medical Services Plan (MSP) database between January 1, 2010 and December 31, 2011 reporting a psychiatric diagnosis (ICD-9 codes 290-319). The exit date is the earliest of December 31, 2012, death, or coverage termination with the Saskatchewan Ministry of Health. Aboriginal identity was indicated by Treaty status and the patient's residence was determined by residence at the index date.

Hospital separation database: The hospital separation database contains information about admission date, discharge date, diagnosis codes and diagnosis type, type of the hospital, designated psychiatry bed, and information about day surgery procedures. The type of hospital was recorded as whether the hospital was categorized as provincial, regional or any other acute care facility in Saskatchewan. The variable psychiatry bed indicated if the patient was admitted to a psychiatric unit if the hospital had one.

Physician services database: The physician services database includes the following

information: date of visit, diagnosis, doctor ID and referring doctor ID. Visit records were created by collapsing service-based claim records submitted to the Medical Services Plan on the following variables: unique identifier, diagnosis, date of service, practitioner number, clinic number, and location of service. That is, all services delivered to a single person by a single physician for the same diagnosis on the same day at the same clinic and same location of service are reduced to a single visit record.

Physician mobility database: Physician mobility information included the doctors ID number, and their medical specialty. Doctor's ID is a sequential unique study reference number assigned by the Ministry of Health and bears no resemblance to a provider's Medical Services Plan (MSP) billing number. The medical specialty of physicians was classified into general practitioner, fee-for-service (FFS) psychiatrist, pediatrician, and any other medical specialists or other non physician providers.

The physician mobility database and physician services database were linked by doctor's ID and then it was linked with the other databases by unique patient ID.

4.2 Outcome Variable

The outcome of interest in this study is the number of hospitalizations among the patients with psychiatric disorders within the first 3, 6, and 9 months of their index dates. We considered those patients who did not have any hospitalizations within 2 years prior to their index dates. The rationale for doing so was to rule out prior hospitalizations as a confounding variable. Patients' visits to general practitioners (GP) and FFS psychiatrists were traced back for two years for mental-health related diagnoses (ICD 10 codes: F00-F99). Demographic characteristics at index date were also extracted. Out of the 200,537 eligible patients, nearly 98% of the patients did not have any hospitalization during these three follow-up periods, as shown in Table 4.1.

4.3 Covariates

The covariates considered include: age of the patients that was calculated from their birth year to the index year and it was grouped by quartiles, registered Indian status (ever/never), gender (male/female), outpatient FFS psychiatrist visit (yes/no) which indicates if there was any FFS psychiatric visits within two years prior to the index date, outpatient general physician (GP) visit (yes/no) which indicates if there was any visit to general physicians within two years prior to their index dates, and disease categories: anxiety, mood disorders, substance use, behavioral disorders and schizophrenia. Table 4.1 presents the descriptive information of the variables, that were classified as clinical characteristics and social-demographic characteristics of the patients over different study periods. Table 4.2 presents summary statistics for the demographic characteristics of the patients before dividing them into three follow up periods. In Table 4.3, some examples of diseases along with the ICD codes are reported for each disease category.

Table 4.1: Descriptive statistics for the number of inpatient hospitalizations

Variable name	3 months		6 months		9 months	
	Yes(n=1266)	No(n=199271)	Yes(n=1864)	No(n=198673)	Yes(n=2300)	No(n=198237)
Clinical characteristics						
Disease category						
Schizophrenia	209(17%)	12364(6%)	305(16%)	12268(6%)	357(16%)	12216(6%)
Anxiety	334(26%)	81840(41%)	475(25%)	81699(41%)	582(25%)	81592(41%)
Behavioral disorder	71(6%)	12517(6%)	111(6%)	12477(6%)	138(6%)	12450(6%)
Mood disorder	481(38%)	60223(30%)	712(38%)	59992(30%)	882(38%)	59822(30%)
Substance use	110(9%)	17833(9%)	174(9%)	17769(9%)	227(10%)	17716(9%)
Others	61(5%)	14494(7%)	87(5%)	14468(7%)	114(5%)	14441(7%)
Outpatient FFS Psychiatrist visit						
Yes	311(25%)	15920(8%)	484(26%)	15747(8%)	575(25%)	15656(8%)
No	955(75%)	183351(92%)	1380(74%)	182926(92%)	1725(75%)	182581(92%)
Outpatient GP visit						
Yes	952(75%)	179788(90%)	1394(75%)	179341(90%)	1752(76%)	178988(90%)
No	314(25%)	19483(10%)	470(25%)	19332(10%)	548(24%)	19249(10%)
Social-demographic characteristics						
Registered Indian						
Ever	217(17%)	21596(11%)	315(17%)	21498(11%)	381(17%)	21432(11%)
Never	1049(83%)	177675(89%)	1549(83%)	177175(89%)	1919(83%)	176805(89%)
Gender						
Male	586(46%)	81478(41%)	849(46%)	81215(41%)	1049(46%)	81015(41%)
Female	680(54%)	117793(59%)	1015(54%)	117458(59%)	1251(54%)	117222(59%)
Age						
(6-29]	466(37%)	51401(26%)	677(36%)	51190(26%)	825(36%)	51042(26%)
(29-45]	268(21%)	49931(25%)	394(22%)	49805 (25%)	509(22%)	49690(25%)
(45-60]	226(18%)	49303(25%)	355(19%)	49174 (25%)	434(19%)	49095(25%)
(60-86]	306(24%)	48636(24%)	438(0.23%)	48504 (24%)	532(23%)	48410(24%)
Residence						
Saskatoon	239(19%)	56860(29%)	358(21%)	56714 (29%)	469(20%)	56630(29%)
Regina	241(19%)	41931(21%)	344(18%)	41828 (21%)	420(18%)	41752(21%)
Lloydminster or others[a]	188(15%)	28239(14%)	293(16%)	28134(14%)	373(16%)	28054(14%)
Rest of Saskatchewan	598(47%)	72241(36%)	842(45%)	71997(36%)	1038(45%)	71801(36%)

a= MooseJaw, North Battleford, Prince Albert, Swift Current, Yorkton

Table 4.2: Descriptive statistics for demographic characteristics of patients with age more than 5 years, before dividing them into different follow-up period

Demographic Characteristics	**Summary statistics**		
	Full data	no hospitalization	hospitalizations
Age (in years)	Mean (45.70) Stdev (20.78)	Mean (45.83) Stdev (20.78)	Mean (43.01) Stdev (22.10)
Gender Male Female	 n=82064 n=118473	 n=80196 n=116160	 n=2061 n=2537
Registered Indian status Never Ever	 n=178724 n=21813	 n=175267 n=21089	 n=3851 n=747

Table 4.3: Categorization of the Diseases

Categories	**Examples of diseases**
Schizophrenia (ICD code: 295, 297, 298)	Simple/acute type of schizophrenia, Paranoid states
Anxiety (ICD code: 300,308, 309)	Phobic states, Neurotic depressive states, Obsessive-compulsive disorders
Behavioral disorder (ICD code: 299, 312-315)	Autism, psychoses with origin specific to childhood, Disturbance of emotions specific to childhood and adolescence.
Mood disorder (ICD code: 296, 311)	Manic-depressive psychosis, Any kind of depressive disorder
Substance use (ICD code: 291, 292, 303)	Alcoholic/drug psychosis, Paranoid or hallucinatory states induced by drugs, Dependent/Non-dependent tobacco/cocaine use disorder.
Others (ICD code: 290, 293, 294)	Dementia, Acute confusional state, Non-alcoholic psychosis, Gender deviations disorders, Predominant disturbance of emotions, Cyclical vomiting/sleep disorder/hair plucking.

4.4 Study Design

For this study, we created a relatively healthy cohort of psychiatric patients, who have not experienced any psychiatric hospitalizations, as an electronic prospective cohort to follow them up. Although some research have found an association between history of previous hospitalizations and repeated hospitalization [100, 101], we implemented wash-out period of two years prior to the index date for following reasons: Hospitalizations due to any psychiatric illness may reflect a severe condition and people with prior hospitalizations would be at greater risk for succeeding ones. Excluding them would be in accord with the principle and practice in analyzing cohort studies. In doing so, all members of our cohort were not previously hospitalized, and therefore a hospitalization would reflect a worsening of their condition.

For this study, we considered three different follow up periods: 3, 6 and 9 months from a study participant's index date. The reason behind choosing three follow-up study periods is to check the consistency of performance of the best model over all the study periods. Again, we wanted to check which variables were consistently significant over the three study periods, which would be an indication that the associations were not spurious or coincidental.

4.5 Results

Poisson, NB, ZIP, ZINB, HP and HNB models were each fit to the data. We also checked possible interaction effects between outpatient psychiatric visit or GP visit and disease category to assess whether outpatient health service use can reduce the hospitalizations for certain type of diseases. No significant interaction effect was identified between outpatient GP visit and disease categories. However, outpatient psychiatric visit and disease category had a significant interaction for the logistic component of the HNB model.

In Tables 4.4-4.6, the results of model comparison are provided on the basis of their AIC and Vuong's test scores. For every study period, HNB had the lowest AIC and -2 log-likelihood. All the models were compared to each other according to Vuong's test results as well. Table 4.4 shows that the result of Vuong's test supports HNB over Poisson as it

produces the score of -5.9581 which is lower than -1.96. The results also support HNB over Poisson, NB, ZIP and HP. Although this test result does not show much difference between the performance of ZINB and HNB but the results of AIC and -2 log-likelihood supports HNB as performing better. The results are also similar for the 6 months and 9 months follow-up period as well and are represented in the tables 4.5 and 4.6. Considering all the results of model comparison tests, we can say that HNB yielded the best model fit among the considered competing models.

To examine how well the models fit the data, the results of RQR tests are provided. The QQ plots of the RQRs, as shown in Figures 4.1-4.3 represent how much each model deviated from normal. Over the three study periods, HNB better satisfied the assumption of normality. The scatter plots of RQRs indicate that the RQRs for the ZINB and HNB are mostly distributed between -4 and 4 as compared with RQRs for other models, which had some RQRs exceeding the limit. Therefore, although all models present some degrees of lack of fit to the data, HNB provides a more satisfactory fit to the data as compared to the other competing models.

In Table 4.7, the estimated area under the ROC curve (AUC) is represented. Each time a single variable was dropped from the full model to assess the predictive power of that single variable on the model. This table shows that our full model has approximately 70% predictive power. Dropping registered Indian status, age, gender, and outpatient GP visit did not affect the predictive power in a significant way. Dropping disease category reduces the AUC to 68% for three and six months follow up period whereas 67% for nine months study period. Dropping residential area of the patients and outpatient FFS psychiatric visit reduces the AUC to 69% .

We further compared the predictive ability of all the models by comparing the observed frequencies versus predicted frequencies for each unique value of the response variable. The results are presented in Table A.1-A.3 for the three study periods, respectively. To ease the comparison, we also present the results in Figures A.1-A.3 to visually compare the predictive ability for all the models. In each of the study periods, the prediction power of HNB outperforms all the other models with predicted frequencies aligning more closely to the observed frequencies as compared to the other competing models.

Table 4.4: Model comparison for the 3 months follow-up study period (p-values for Vuong's test are given in the parentheses).

Model	AIC	-2 Log likelihood	Vuong's test				
			NB	ZIP	ZINB	HP	HNB
Poisson	16839	16807	-5.812 (<0.001)	-6.164 (<0.001)	-6.035 (<0.001)	-6.004 (0.009)	-5.958 (<0.001)
NB	15689	15655		-1.477 (<0.069)	-4.852 (<0.001)	-0.764 (0.222)	-4.236 (<0.001)
ZIP	15664	15590			-3.823 (<0.001)	2.346 (0.009)	-3.269 (<0.001)
ZINB	15572	15496				4.615 (<0.001)	-0.178 (0.4292)
HP	15644	15570					-4.444 (<0.001)
HNB	15534	15458					

Table 4.5: Model comparison for the 6 months follow-up study period (p-values for Vuong's test are given in the parentheses).

Model	AIC	-2 Log likelihood	Vuong's test				
			NB	ZIP	ZINB	HP	HNB
Poisson	24140	24108	-7.116 (<0.001)	-7.993 (<0.001)	-7.473 (<0.001)	-7.712 (<0.001)	-7.268 (<0.001)
NB	22092	22058		-0.465 (<0.320)	-6.032 (<0.001)	0.523 (0.300)	-4.832 (<0.001)
ZIP	22084	22010			-3.572 (<0.001)	3.741 (0.009)	-2.815 (0.002)
ZINB	21893	21817				4.737 (<0.001)	-0.967 (0.166)
HP	22073	21999					-4.013 (<0.001)
HNB	21829	21753					

Table 4.6: Model comparison for the 9 months follow-up study period (p-values for Vuong's test are given in the parentheses).

Model	AIC	-2 Log likelihood	Vuong's test				
			NB	ZIP	ZINB	HP	HNB
Poisson	30179	30147	-8.269 (<0.001)	-9.527 (<0.001)	-8.726 (<0.001)	-9.527 (<0.001)	-8.491 (<0.001)
NB	26863	26829		0.715 (<0.237)	-6.431 (<0.001)	1.704 (0.044)	-5.439 (<0.001)
ZIP	26948	26874			-4.220 (<0.001)	4.410 (<0.001)	-3.438 (<0.001)
ZINB	26578	26502				5.310 (<0.001)	-1.355 (0.087)
HP	26984	26910					-4.539 (<0.001)
HNB	26555	26479					

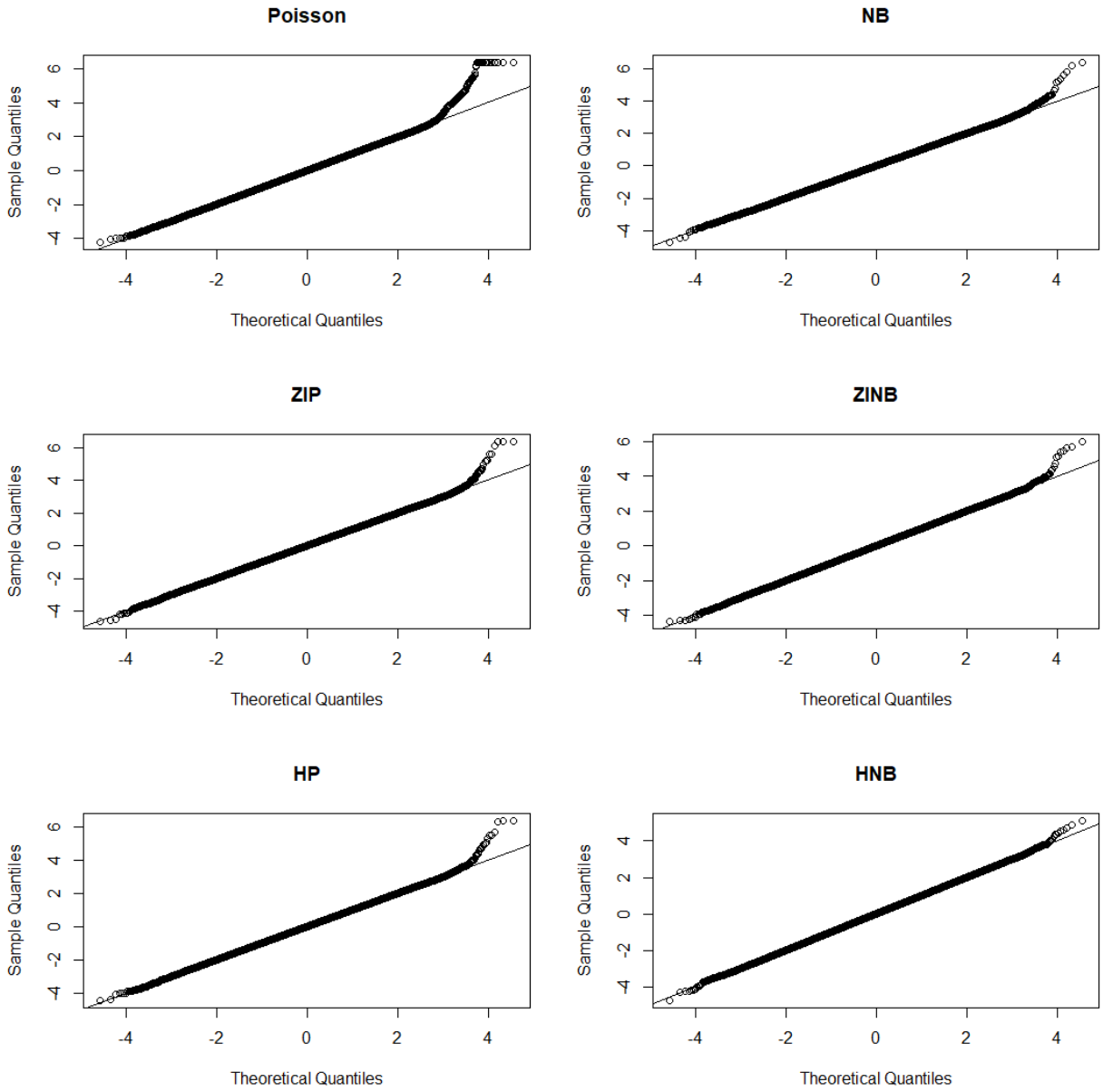


Figure 4.1: QQ normality plots of RQRs for all the competing models for the 3 months follow-up study period.

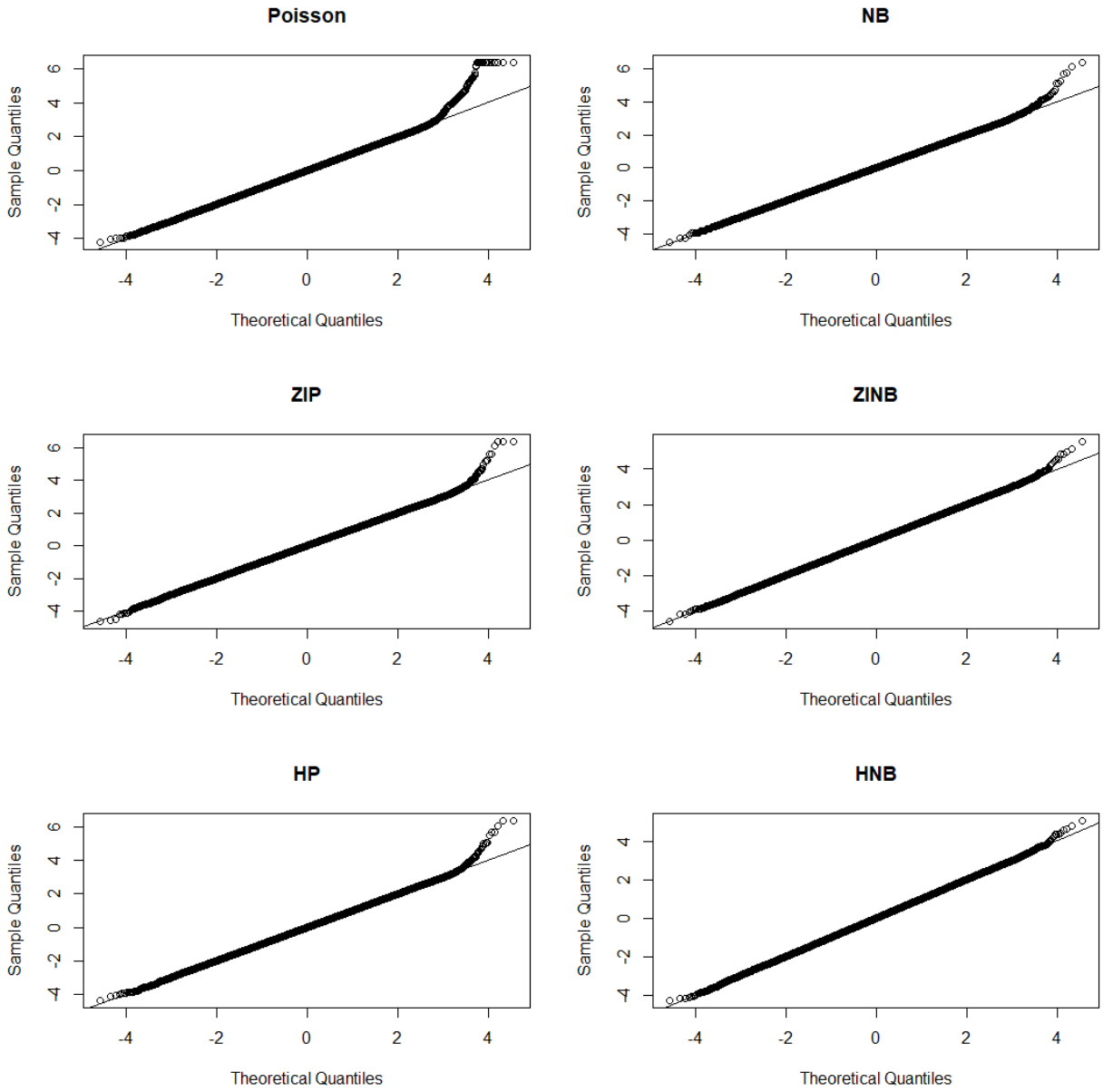


Figure 4.2: QQ normality plots of RQRs for all the competing models for the 6 months follow-up study period.

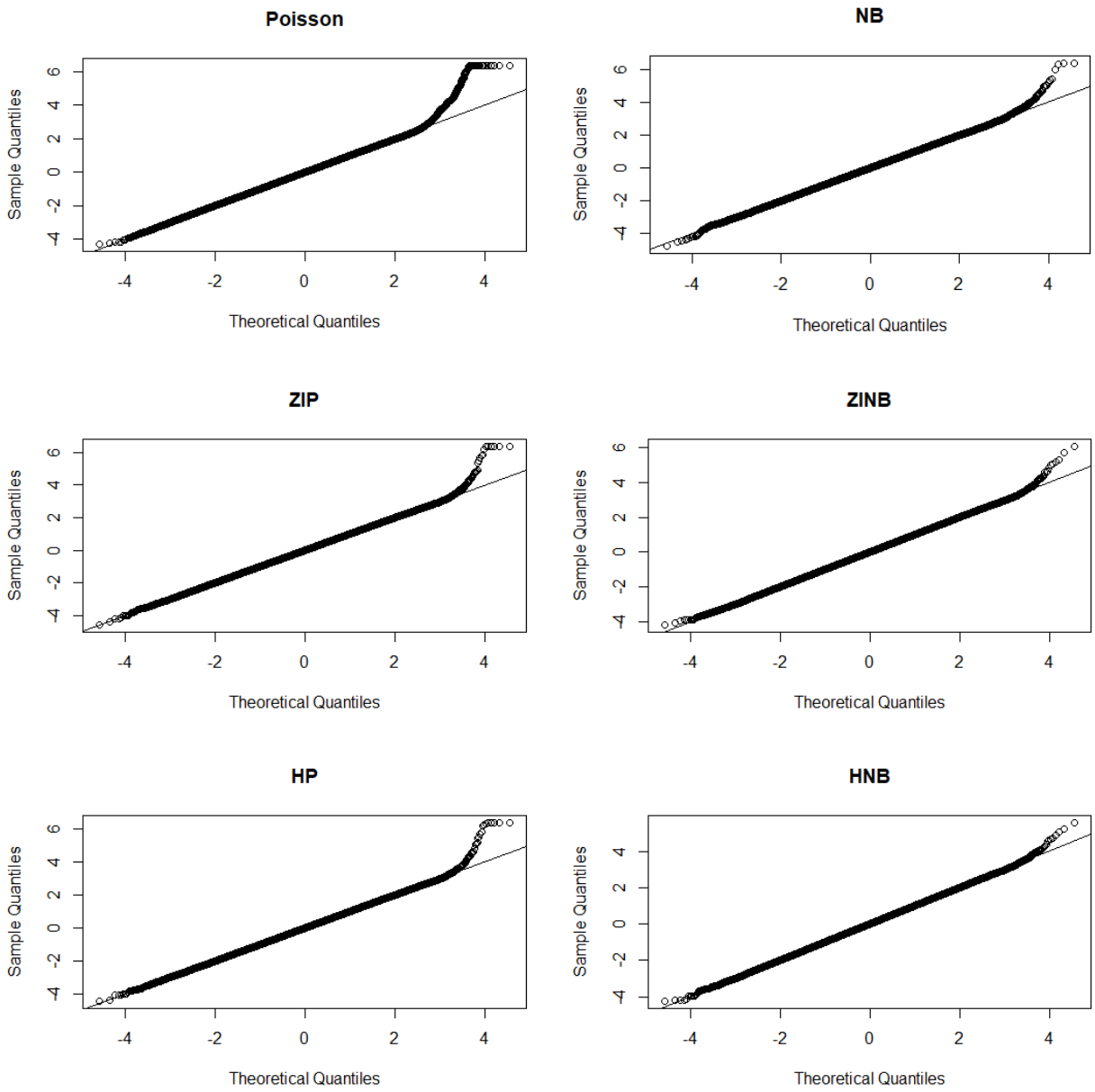
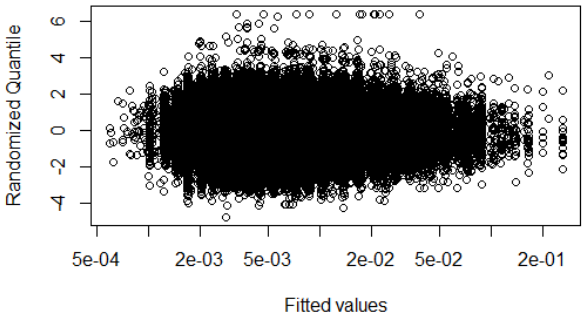
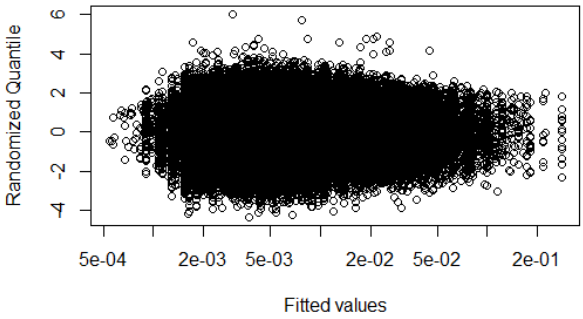


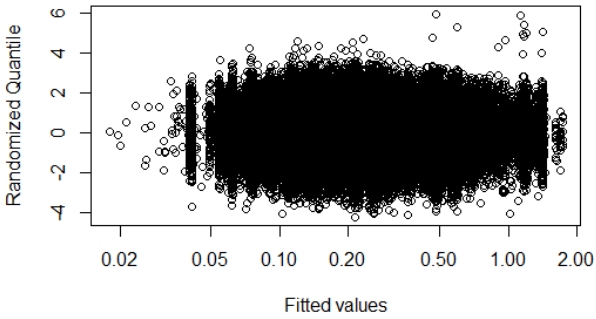
Figure 4.3: QQ normality plots of RQRs for all the competing models for the 9 months follow-up study period.



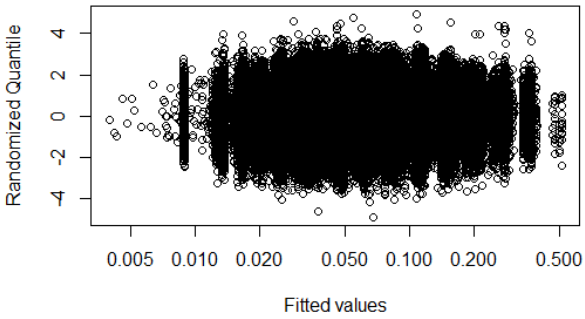
(a) Poisson



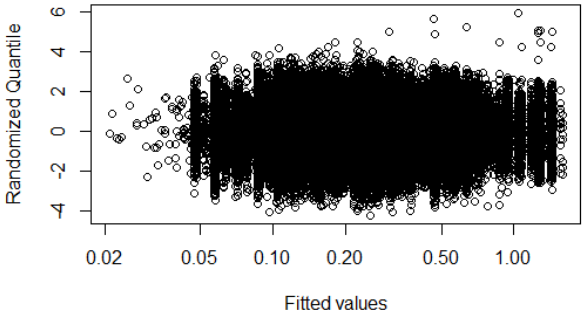
(b) NB



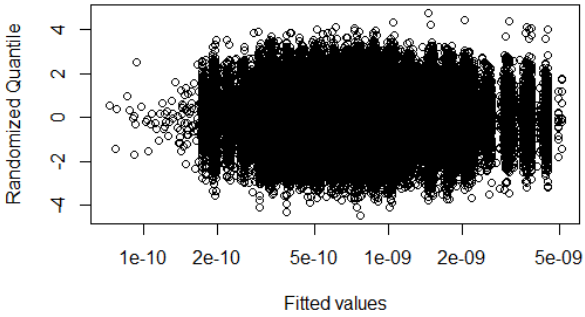
(c) ZIP



(d) ZINB

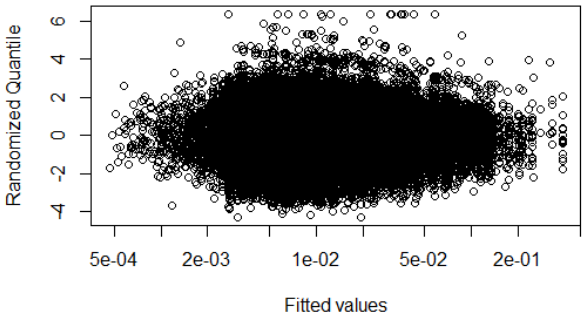


(e) HP

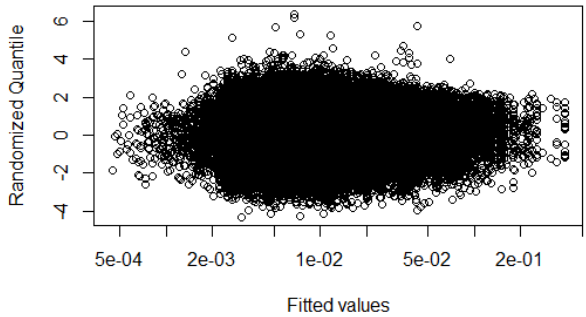


(f) HNB

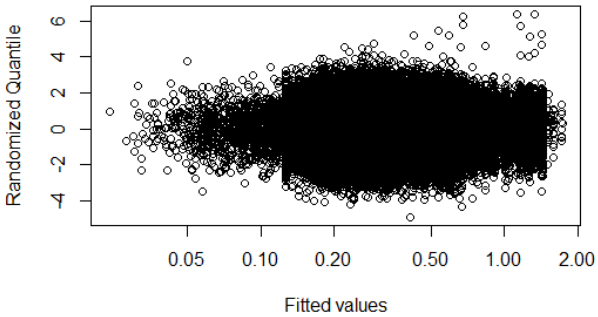
Figure 4.4: Scatter plots of RQRs for all the competing models for the 3 months follow-up study period.



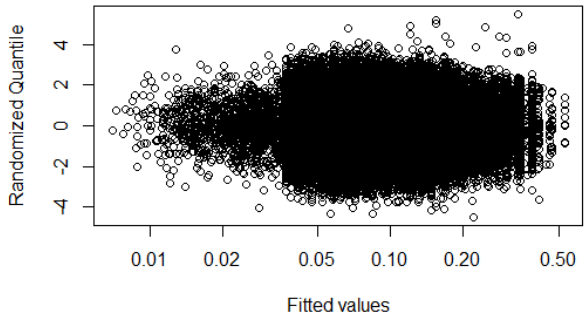
(a) Poisson



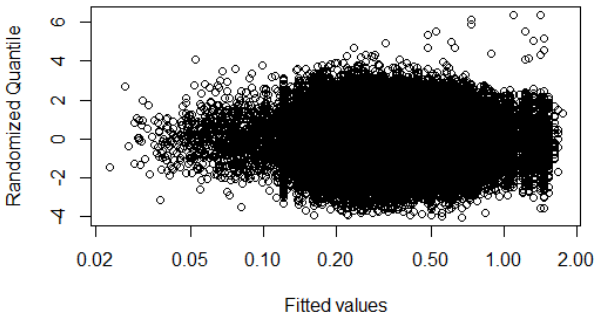
(b) NB



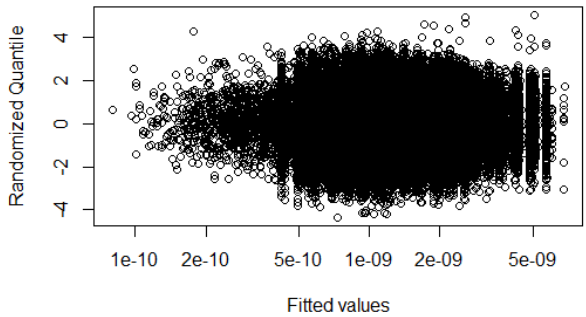
(c) ZIP



(d) ZINB

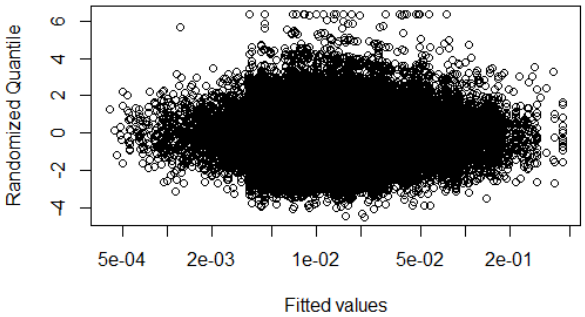


(e) HP

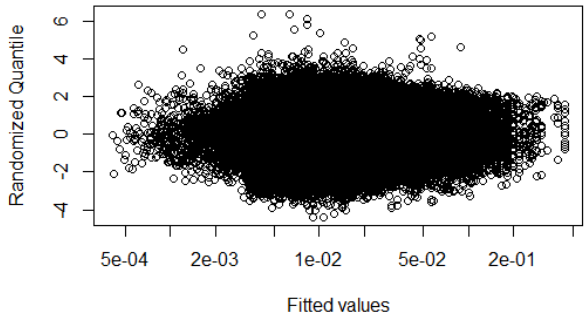


(f) HNB

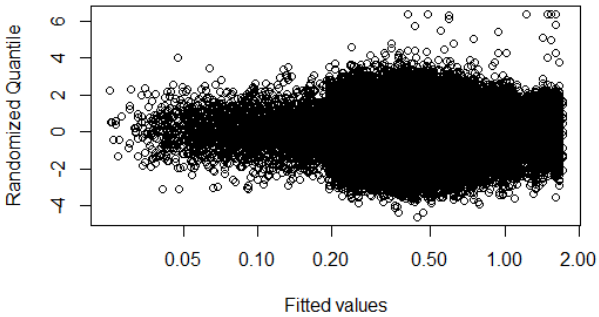
Figure 4.5: Scatter plots of RQRs for all the competing models for the 6 months follow-up study period



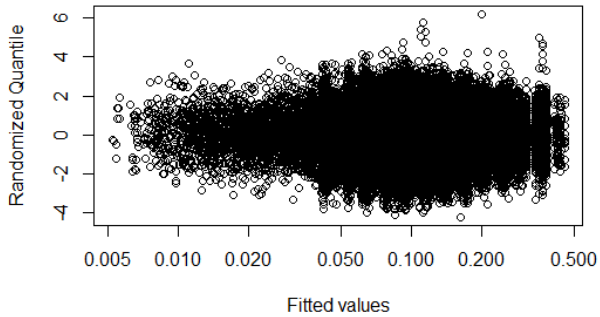
(a) Poisson



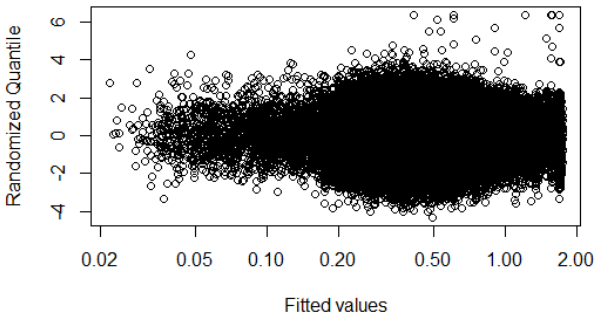
(b) NB



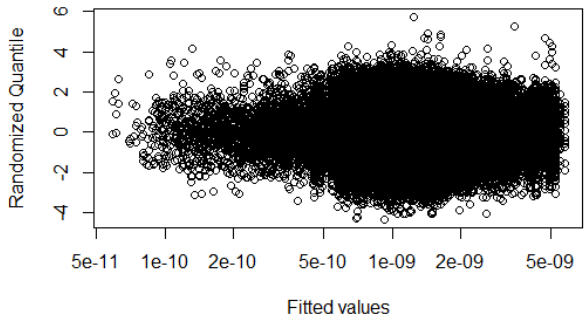
(c) ZIP



(d) ZINB



(e) HP



(f) HNB

Figure 4.6: Scatter plots of RQRs for all the competing models for the 9 months follow-up study period.

Table 4.7: Area under ROC curve (AUC) of the logistic component of the HNB model for the 3, 6, and 9 months follow-up periods.

	AUC		
	3 Months	6 Months	9 months
Full Model	0.7143	0.7128	0.6941
Dropped covariate			
Gender	0.7125	0.7113	0.6928
Outpatient GP visit	0.7119	0.7107	0.6924
Registered Indian Status	0.7114	0.7109	0.6933
Age	0.7084	0.707	0.6877
Outpatient Psychiatric visit	0.6979	0.6923	0.6747
Residence	0.6903	0.69	0.6743
Disease Category	0.6894	0.6844	0.6645

Tables 4.8-4.10 present the estimated regression coefficients of the best model considered, i.e., HNB model, for the 3, 6 and 9 months follow-up periods, respectively. In this study, the factors that explain hospitalization were grouped into: clinical factors and socio-demographic factors. The results consist of two separate parts: one is for the positive counts or the number of hospitalizations for those who had at least one hospitalization and the other one is for hospitalization vs no hospitalization. The results are also shown in Figures 4.7-4.9 in terms of 95% confidence interval for odds ratio and risk ratio for the main effects. For the logistic model, Table 4.8 shows that for patients who are registered Indian, the odds of being hospitalized due to psychiatric disorder is 1.59 (95% CI: 1.36, 1.86) times higher than those who are not registered Indian. The odds of being hospitalized for those patients who visited GP prior to their index dates are 0.63 (95% CI: 0.53, 0.75) times lower than those patients who did not visit GP for psychiatric concerns. Patients from Lloydminster, Regina and the rest of Saskatchewan are respectively 1.93 (95% CI: 1.58, 2.34), 1.64 (95% CI: 1.37, 1.97) and 2.38 (95% CI: 2.03, 2.78) times more likely to be hospitalized than patients from Saskatoon. The odds of getting hospitalized for the patients aged between 6 to 29 years and 30 to 45 years old are 2.15 (95% CI: 1.81, 2.56) and 1.22 (95%CI: 1.01, 1.46) times higher than those who are aged between 61 to 86. Males are 1.22 (95% CI: 1.09, 1.37) times more likely to be hospitalized than females. The interaction effects between disease category and outpatient psychiatric visits over 3, 6 and 9 months follow-up periods are presented in the Tables 4.11-4.13 and Figure 4.10 in terms of odds ratio and 95% confidence interval for odds

ratio. Table 4.11 shows that patients who suffers from Schizophrenia and have visited any FFS psychiatrist are 2.005 (95% CI: 1.355, 2.965) times more likely to be hospitalized comparing to those patients who suffers from "others" disease category. Again, those patients who suffers from Schizophrenia and did not visit any FFS psychiatrist are 4.9 (95% CI: 3.806, 6.473) times more likely to be hospitalized comparing to those who suffers from "others" category of diseases. This indicates that for patients suffering from Schizophrenia previous visits to any FFS psychiatrist can play as a protective factor against hospitalization in comparison to the "others" disease category. For the patients who suffered from Anxiety, Mental disorder due to substance use and Mood disorder have higher odds i.e, respectively 1.205 (95% CI: 0.948, 1.531), 1.670 (95% CI: 1.278, 2.181), 2.370 (95% CI: 1.871, 3.001) times more likely to be hospitalized comparing to the patients who suffered from "others" category of diseases. Whereas, the odds of getting hospitalized for the patients suffering from Behavioral disorder is 0.877 (95% CI: 0.636, 1.209) times lower than the patients suffering from "others" category of diseases. The results are quite similar for the rest of the study periods.

The count part of the model shows that the rate for patients who previously visited psychiatrists are 1.99 (95% CI: 1.33, 2.97) times more likely to have multiple hospitalizations comparing to those who did not visit any psychiatrist. The risk of multiple hospitalizations is 0.58 (95% CI: 0.31, 1.08), 0.37 (95% CI: 0.20,0.68) and 0.84 (95% CI: 0.53, 1.31) times lower for the patients from Lloydminster, Regina and Rest of Saskatchewan respectively than the patients from Saskatoon. Younger patients aged between 6 to 29 years old have 0.61 (95% CI: 0.38, 1.00) times lower risk of multiple hospitalizations than the patients aged between 61 to 86. No significant effects for the disease category were observed for the count part of the HNB model. The results are quite similar for the 6 months and 9 months study periods as well.

Table 4.8: Parameter estimates of the best fitted model (HNB model) for the 3 months follow up period (n=200,537). OR: odds ratio, RR: risk ratio.

logit$[P(Y_i > 0)]$	$\hat{\tilde{\beta}}$	SE	P-value	OR	95% CI of OR	
					Lower	Upper
Registered Indian (Yes vs No)	0.463	0.080	<0.001*	1.59	1.36	1.86
GP visit (Yes vs No)	-0.468	0.089	<0.001*	0.63	0.53	0.75
Lloydminster vs Saskatoon	0.655	0.099	<0.001*	1.93	1.58	2.34
Regina vs Saskatoon	0.866	0.093	<0.001*	1.64	1.37	1.97
Rest of Saskatchewan vs Saskatoon	0.866	0.080	<0.001*	2.38	2.03	2.78
Age [6,29] vs Age [61,86]	0.766	0.079	<0.001*	2.15	1.81	2.56
Age [30,45] vs Age [61,86]	0.201	0.089	0.041*	1.22	1.01	1.46
Age [46,60] vs Age [61,86]	-0.01	0.101	0.941	0.99	0.82	1.20
Gender (Male vs Female)	0.198	0.058	<0.001*	1.22	1.09	1.37
log$[E(Y_i \mid Y_i > 0)]$	$\hat{\beta}$	SE	P-value	RR	95% CI of RR	
					Lower	Upper
Psychiatric visit (Yes vs No)	0.689	0.203	<0.001*	1.99	1.33	2.97
Lloydminster vs Saskatoon	-0.541	0.319	0.090	0.58	0.31	1.08
Regina vs Saskatoon	-0.991	0.314	0.001*	0.37	0.20	0.68
Rest of Saskatchewan vs Saskatoon	-0.174	0.227	0.444	0.84	0.53	1.31
Age [6,29] vs Age [61,86]	-0.481	0.246	0.050*	0.61	0.38	1.00
Age [30,45] vs Age [61,86]	0.040	0.251	0.862	1.04	0.63	1.71
Age [46,60] vs Age [61,86]	0.171	0.260	0.521	1.18	0.70	1.97
Anxiety vs Others	0.067	0.484	0.888	1.07	0.41	2.76
Behavioral disorder vs Others	0.362	0.603	0.548	1.43	0.43	4.69
Substance use vs Others	0.104	0.552	0.849	1.11	0.37	3.28
Mood disorder vs Others	0.749	0.459	0.102	2.11	0.85	5.20
Schizophrenia vs Others	0.245	0.496	0.621	1.27	0.48	3.38

Psychiatric visit: Outpatient Psychiatric visit prior to index date.
GP visit: Outpatient GP visit prior to index date.

Table 4.9: Parameter estimates of the best fitted model (HNB model) for 6 months follow up period (n=200,537). OR: odds ratio, RR: risk ratio.

logit$[P(Y_i > 0)]$	$\hat{\hat{\beta}}$	SE	P-value	OR	95% CI of OR	
					Lower	Upper
Registered Indian (Yes vs No)	0.445	0.066	<0.001*	1.56	1.37	1.78
GP visit (Yes vs No)	-0.392	0.073	<0.001*	0.68	0.58	0.78
Lloydminster vs Saskatoon	0.640	0.079	<0.001*	1.90	1.62	2.22
Regina vs Saskatoon	0.385	0.076	<0.001*	1.47	1.27	1.71
Rest of Saskatchewan vs Saskatoon	0.755	0.064	<0.001*	2.13	1.88	2.42
Age [6,29] vs Age [61,86]	0.776	0.074	<0.001*	2.17	1.88	2.52
Age [30,45] vs Age [61,86]	0.222	0.078	<0.001*	1.25	1.07	1.46
Age [46,60] vs Age [61,86]	0.098	0.079	<0.001*	1.10	0.94	1.29
Gender (Male vs Female)	0.149	0.048	<0.001*	1.16	1.06	1.28
log$[E(Y_i \mid Y_i > 0)]$	$\hat{\beta}$	SE	P-value	RR	95% CI of RR	
					Lower	Upper
Psychiatric visit (Yes vs No)	0.315	0.204	0.122	1.37	0.92	2.05
GP visit (Yes vs No)	-0.464	0.203	0.022*	0.63	0.42	0.94
Lloydminster vs Saskatoon	-0.289	0.221	0.190	0.75	0.48	1.15
Regina vs Saskatoon	-0.631	0.218	0.003*	0.53	0.35	0.82
Rest of Saskatchewan vs Saskatoon	-0.130	0.168	0.436	0.88	0.63	1.22
Age [6,29] vs Age [61,86]	-0.251	0.178	0.174	0.78	0.54	1.11
Age [30,45] vs Age [61,86]	-0.091	0.183	0.640	0.91	0.62	1.34
Age [46,60] vs Age [61,86]	0.002	0.194	0.170	1.00	0.68	1.48
Gender (Male vs Female)	0.147	0.129	0.253	1.16	0.90	1.49
Anxiety vs Others	-0.187	0.335	0.575	0.83	0.43	1.60
Behavioral disorder vs Others	-0.142	0.417	0.733	0.87	0.38	1.97
Substance use vs Others	0.136	0.371	0.712	1.15	0.55	2.38
Mood disorder vs Others	0.501	0.317	0.114	1.65	0.89	3.07
Schizophrenia vs Others	-0.204	0.346	0.555	0.81	0.41	1.61

Psychiatric visit: Outpatient Psychiatric visit prior to index date.
GP visit: Outpatient GP visit prior to index date.

Table 4.10: Parameter estimates of the best fitted model (HNB model) for the 9 months follow up period (n=200,537). OR: odds ratio, RR: risk ratio.

logit$[P(Y_i > 0)]$	$\hat{\tilde{\beta}}$	SE	P-value	OR	95% CI of OR	
					Lower	Upper
Registered Indian (Yes vs No)	0.401	0.060	<0.001*	1.49	1.33	1.68
GP visit (Yes vs No)	-0.345	0.067	<0.001*	0.71	0.62	0.81
Lloydminster vs Saskatoon	0.681	0.071	<0.001*	1.98	1.72	2.28
Regina vs Saskatoon	0.381	0.069	<0.001*	1.46	1.28	1.68
Rest of Saskatchewan vs Saskatoon	0.771	0.058	<0.001*	2.16	1.93	2.43
Age [6,29] vs Age [61,86]	0.780	0.071	<0.001*	2.19	1.92	2.50
Age [30,45] vs Age [61,86]	0.290	0.071	<0.001*	1.33	1.16	1.53
Age [46,60] vs Age [61,86]	0.111	0.069	0.120	1.12	0.97	1.29
Gender (Male vs Female)	0.146	0.043	<0.001*	1.16	1.06	1.26
log$[E(Y_i \mid Y_i > 0)]$	$\hat{\beta}$	SE	P-value	RR	95% CI of RR	
					Lower	Upper
Psychiatric visit (Yes vs No)	0.483	0.170	<0.001*	1.62	1.16	2.27
GP visit (Yes vs No)	-0.411	0.171	0.016*	0.66	0.47	0.93
Lloydminster vs Saskatoon	-0.340	0.181	0.060	0.71	0.50	1.02
Regina vs Saskatoon	-0.652	0.182	<0.001*	0.52	0.36	0.74
Rest of Saskatchewan vs Saskatoon	-0.208	0.141	0.142	0.81	0.62	1.07
Age [6,29] vs Age [61,86]	0.031	0.157	0.841	1.03	0.76	1.40
Age [30,45] vs Age [61,86]	0.167	0.165	0.313	1.18	0.85	1.64
Age [46,60] vs Age [61,86]	0.199	0.171	0.245	1.22	0.87	1.71
Anxiety vs Others	-0.037	0.268	0.889	0.96	0.57	1.63
Behavioral disorder vs Others	-0.163	0.333	0.624	0.85	0.44	1.63
Substance use vs Others	-0.047	0.303	0.874	0.95	0.53	1.73
Mood disorder vs Others	0.293	0.255	0.250	1.34	0.81	2.21
Schizophrenia vs Others	-0.316	0.284	0.266	0.73	0.42	1.27

Psychiatric visit: Outpatient Psychiatric visit prior to index date.
GP visit: Outpatient GP visit prior to index date.

Table 4.11: Odds ratio for the interaction effects between disease category and FFS psychiatric visits over 3 months.

FFS psychiatric visits	**Disease category**	**Odds ratio**		
		Estimate	95% Lower confidence limit	95% Upper confidence limit
Yes	Anxiety vs Others	0.647	0.434	0.963
	Behavioral disorder vs Others	0.392	0.256	0.602
	Substance use vs Others	1.193	0.741	1.922
	Mood disorder vs Others	0.917	0.635	1.324
	Schizophrenia vs Others	2.005	1.355	2.965
No	Anxiety vs Others	1.205	0.948	1.531
	Behavioral disorder vs Others	0.877	0.636	1.209
	Substance use vs Others	1.670	1.278	2.181
	Mood disorder vs Others	2.370	1.871	3.001
	Schizophrenia vs Others	4.963	3.806	6.473

Table 4.12: Odds ratio for the interaction effects between disease category and FFS psychiatric visits over 6 months.

FFS psychiatric visits	**Disease category**	**Odds ratio**		
		Estimate	95% Lower confidence limit	95% Upper confidence limit
Yes	Anxiety vs Others	0.698	0.450	1.084
	Behavioral disorder vs Others	0.442	0.277	0.704
	Substance use vs Others	1.184	0.696	2.015
	Mood disorder vs Others	0.968	0.644	1.455
	Schizophrenia vs Others	2.110	1.369	3.253
No	Anxiety vs Others	1.316	0.998	1.734
	Behavioral disorder vs Others	0.833	0.574	1.211
	Substance use vs Others	1.703	1.252	2.316
	Mood disorder vs Others	2.560	1.950	3.359
	Schizophrenia vs Others	5.754	4.257	7.776

Table 4.13: Odds ratio for the interaction effects between disease category and FFS psychiatric visits over 9 months.

FFS psychiatric visits	**Disease category**	**Odds ratio**		
		Estimate	95% Lower confidence limit	95% Upper confidence limit
Yes	Anxiety vs Others	0.674	0.434	0.963
	Behavioral disorder vs Others	0.392	0.256	0.602
	Substance use vs Others	1.193	0.741	1.922
	Mood disorder vs Others	0.917	0.635	1.324
	Schizophrenia vs Others	2.005	1.355	2.965
No	Anxiety vs Others	1.205	0.948	1.531
	Behavioral disorder vs Others	0.877	0.636	1.209
	Substance use vs Others	1.670	1.278	2.181
	Mood disorder vs Others	2.370	1.871	3.001
	Schizophrenia vs Others	4.963	3.806	6.473

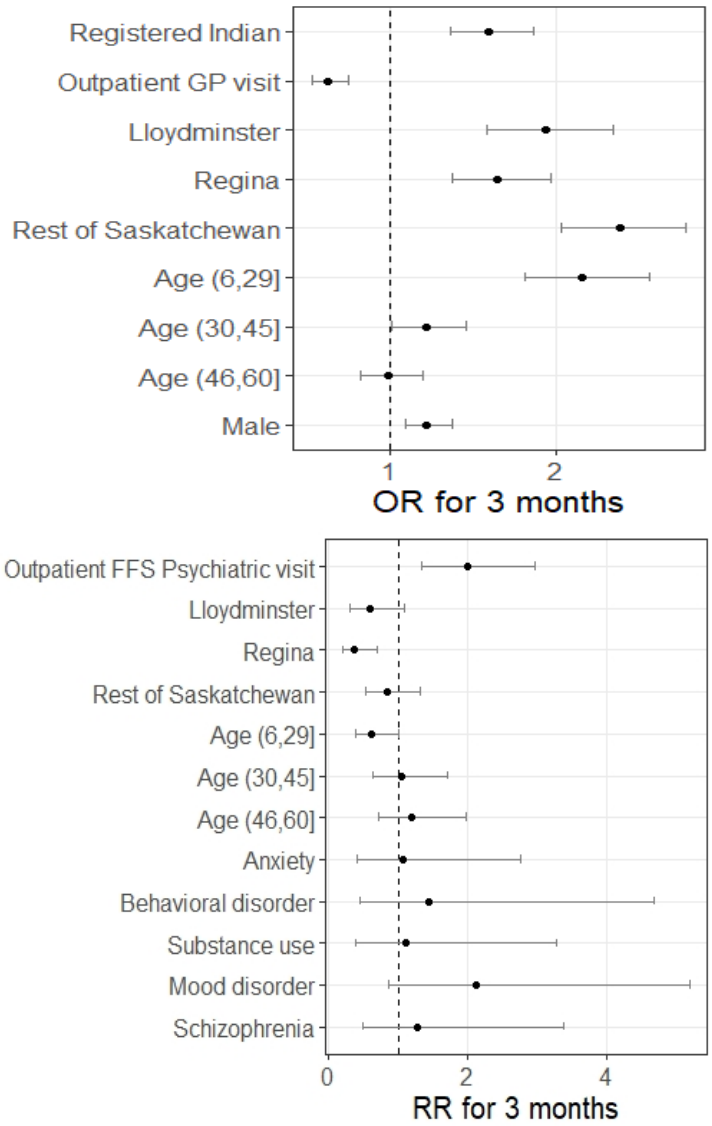


Figure 4.7: Odds ratio (OR) and Risk ratio (RR) plots for the best fitted model, i.e, HNB for the 3 months follow-up study.

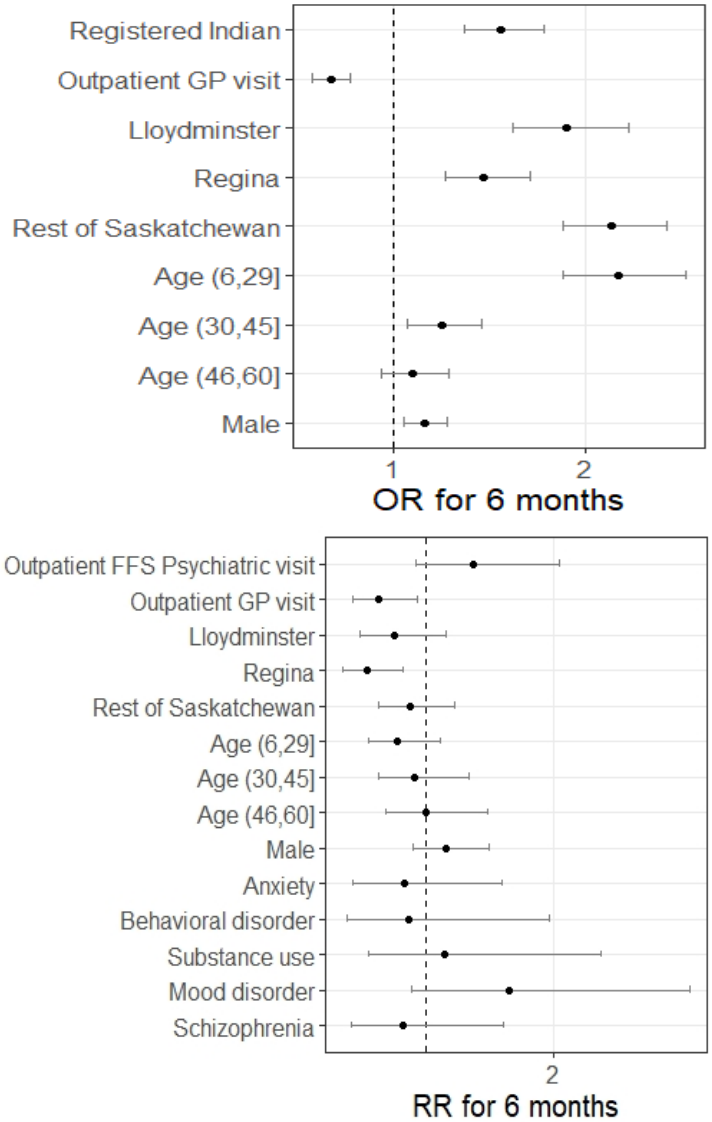


Figure 4.8: Odds ratio (OR) and Risk ratio (RR) plots for the best fitted model, i.e, HNB for the 6 months follow-up study.

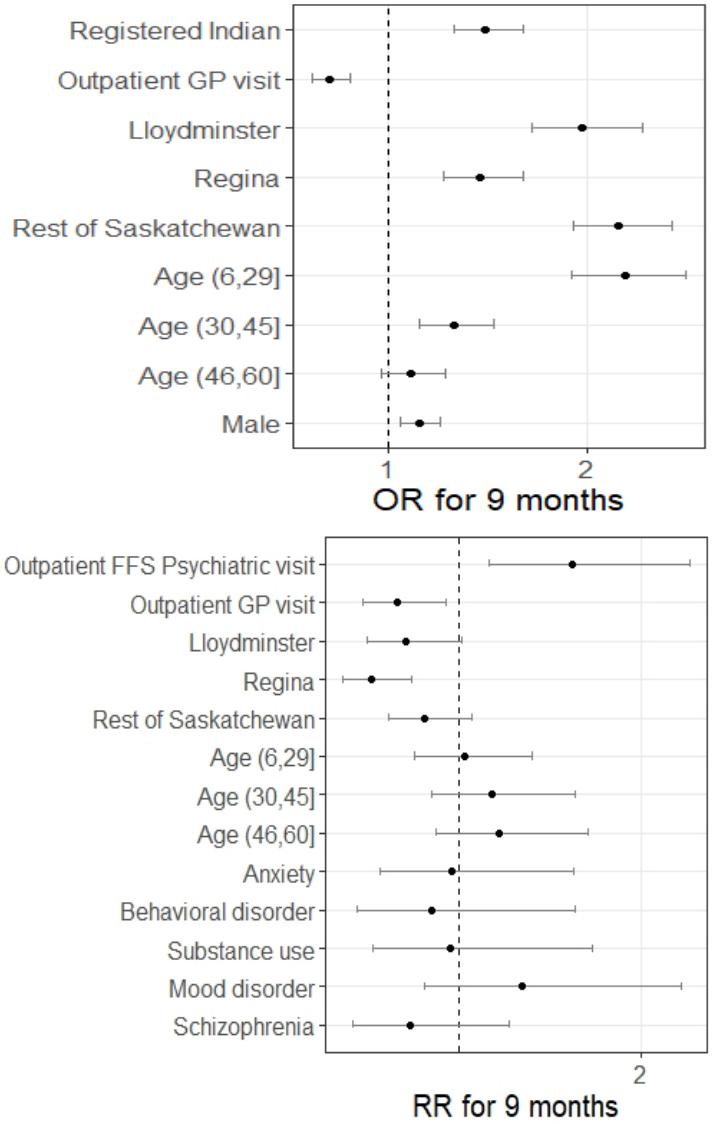


Figure 4.9: Odds ratio (OR) and Risk ratio (RR) plots for the best fitted model, i.e, HNB for the 9 months follow-up study.

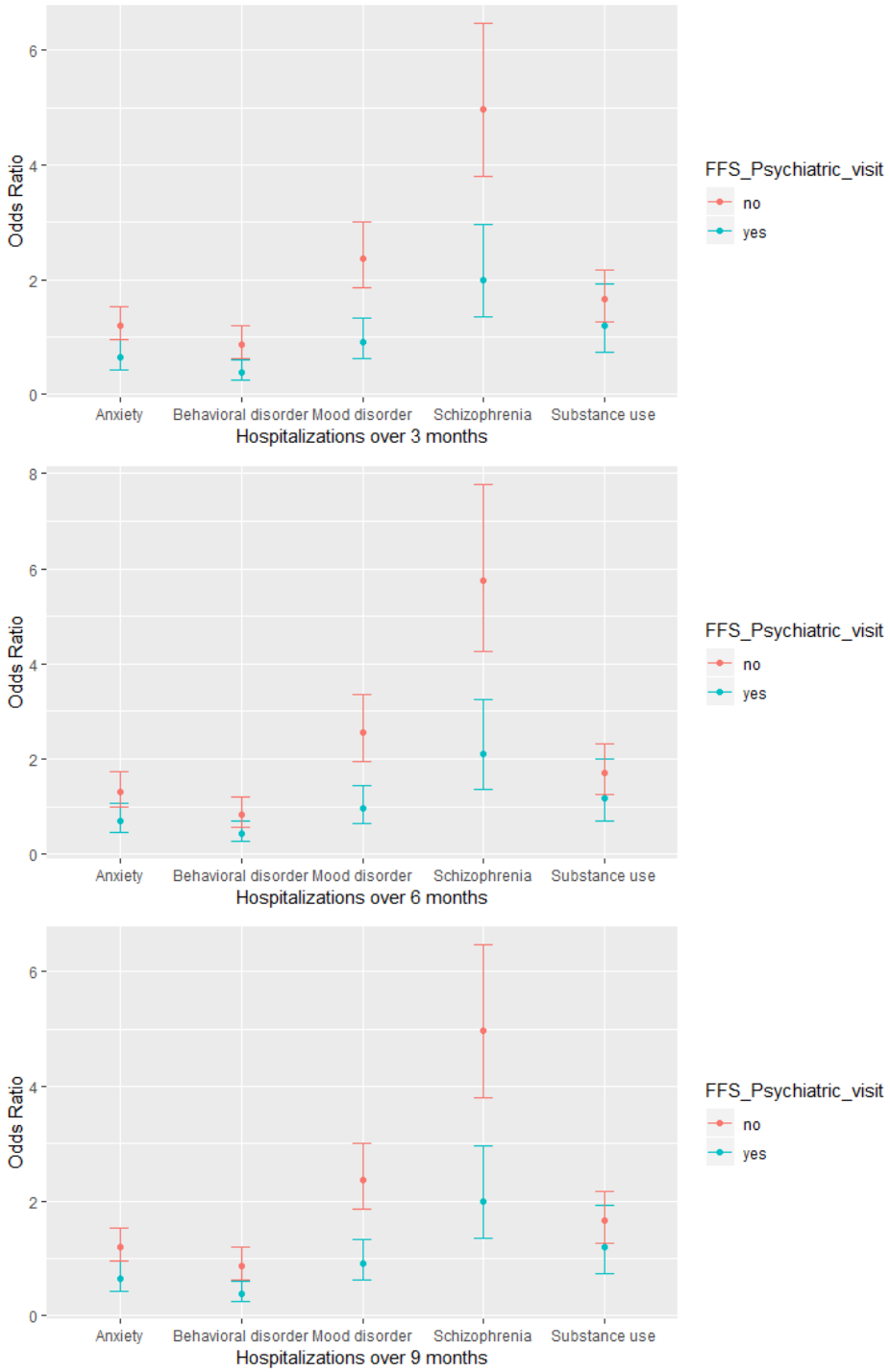


Figure 4.10: Odds ratios of being hospitalized for disease categories (Anxiety, Behavioral disorder, Mental disorder due to substance use, Mood disorder and Schizophrenia vs others) by outpatient FFS psychiatric visits status (Yes. vs. No) for the logistic component of the best-fitting model, i.e., HNB model at the 3, 6 and 9 months follow-ups.

5. Discussion

Our results (Tables 4.8-4.10) indicate that the odds of having at least one hospitalization vs. no hospitalization for mental disorders were significantly higher among Aboriginal than non-Aboriginal people, but no significant difference was detected in the hospital readmission rate between Aboriginal vs. non-Aboriginal people based on the conditional counts component in the HNB models. This demonstrates the potential advantage of HNB models to provide a more precise interpretation of the data when the process that generates zero values differs from the process that generates positive counts.

Previous literature suggested a number of factors which may contribute to the higher hospitalization rates for mental or behavioural disorders among Aboriginal people. Those factors include the trauma and disempowerment caused by residential schools; the forced relocation of communities; and the forced removal of children away from their families. Those issues may have placed Aboriginal people at a higher risk of mental illnesses [102–106] such as depression and psychological distress [107–109]. Inequalities in the social determinants of health may also influence hospitalization rate disparities. In many First Nations communities, educational and employment opportunities are limited and the prevalence of low income is high [110]. It is also possible that they may encounter barriers when they seek primary health care [108, 109, 111] or perceive discrimination as patients [112].

Over the past decade, the prevalence of mental health diagnoses has been rising among young patients seeking acute medical care [113]. A recent comprehensive review of the field of child psychiatric epidemiology [114] noted that the number of observations with mental health issues in community surveys of children and adolescents has risen from 10,000 in studies published between 1980 and 1993 to nearly 40,000 from 21 studies published between 1993 and 2002 [115]. The results of these studies indicate that about one out of every three to four youths is estimated to meet lifetime criteria for a Diagnostic and Statistical Manual of Mental Disorders (DSM) mental disorder [114]. However, a small proportion of these youth actually have sufficiently severe distress or impairment to warrant intervention [116]. About one out

of every ten youths is estimated to meet the Substance Abuse and Mental Health Services Administration (SAMHSA) criteria for a Serious Emotional Disturbance (SED) [115, 116].

Our results support this finding, as the logistic components of the HNB models indicate that age is negatively associated with propensity of hospitalization, i.e., younger people ages from 6 to 29 are more likely to be hospitalized for mental health concerns (Figures 4.7-4.9). Nevertheless, as shown in the results for the counts components of the HNB models, no significant association between the number of repeated hospitalizations and age was identified for those patients who had at least one hospitalization during the study period for mental health concerns. We speculate that younger people are more likely to be hospitalized for urgent help for mental illnesses, which might imply that young people who were dealing with serious anxiety or depression had lack of access to counseling services or outpatient FFS psychiatric care. This suggests that younger population are a priority population for the development of a standard approach to ensure adequate resources for this population with mental health conditions.

According to World Health Organization (WHO) [51], sex/gender differences are common in the rates of common mental disorders - depression, anxiety and somatic complaints. These disorders, which have higher prevalence among women, affect approximately 1 in 3 people in the community and constitute a serious public health problem. Some studies reported that although females have a higher prevalence rate, burden of illness, and likelihood of seeking outpatient treatment for psychiatric disorders; they are less likely than males to receive formal mental health care services, and more likely to receive pharmacological prescriptions from primary care providers [117–120]. Some of the possible reasons of the gender differences in access to mental health care may be because of women's autonomy, child bearing responsibilities or health literacy regarding psychiatric illness. Our results based on the logistic regression part of the HNB model indicate that males are more likely to be hospitalized. This result is consistent over the three study periods. On the other hand, for the counts regression component of the HNB model, gender did not play any significant role over three follow-up periods. Further investigation is needed to understand the inconsistency of our finding with the literature.

In several previous research studies, significant area differences in readmission rates were

found. Studies including area/region/country variables, or comparing different areas, may capture system level differences related both to regulation, financing and governance, capacity, organization and structure, as well as environmental factors, but do not directly investigate specific system variables. However, some studies that include area level variables provide an explicit motivation/discussion that relates to health care system differences. Lower readmission rates in urban regions were found in two studies [52, 53]. Ruesch et al.[52] also found a positive association between readmission rates and population density, i.e. combined, the results of the two variables seem to capture non-linearity in the effect of density. Other studies, however, did not find any association between population density [121, 122] or distance to services [123] and readmission rates. In our study, the results from count regression show that the readmission rate in Regina is lower than Saskatoon, which could be possibly due to the difference in population density in those areas. There could be some other underlying reasons as well, like distance to the nearest inpatient service, availability of community health services and factors that are likely to affect service use and aggregate service needs. However, based on the logistic component of the model, patients of Saskatoon are less likely to be hospitalized due to their mental conditions compared to Regina, Lloydminster and the rest of Saskatchewan.

For the primary variables of interest for this study, i.e., outpatient psychiatric or general physician mental health care, our results based on the counts regression component of the HNB models indicate that visiting a general physician prior to the index date protects patients from having multiple hospitalizations. The results from logistic component show that visiting a general physician in the two years prior to the index visit plays a protective role in case of hospitalization. One possible interpretation of these results could be that visits to general physician may reflect a clinical assessment of lower risk or severity as compared with patients referred to acute services. Referral to more specialized services (e.g. FFS psychiatrist vs. other mental health professional, community mental health teams vs. outpatient follow-up) also seems to increase the readmission risk. This may indicate that patients are not seen by psychiatrists until they are very seriously ill. It is assumed that people who are referred to a psychiatrist usually have a more serious condition that is better handled by a specialist in mental health, rather than a general physician. The association between visit to any FFS

psychiatrist and higher admission rate could also indicate that those patients were in the psychiatric waiting list for sometime but as they had a severe issue, had to end up in a hospital. It was found in a study of hospitalization due to mental health diseases [124] that, most patients had a diagnosis of psychosis (60%) and the most common reason for admission was for risk containment. This might reflect the fact that public mental health services mainly provide services to people with severe mental illness and a psychiatric admission is often preceded by behavioural disturbance that could not be managed in the community. Community based resources of psychiatric services are more resourced and skilled in providing services to these patients in the modern era of mental health service delivery in developed countries [125–127].

The "others" category in the disease type consists of diseases like: dementias, personality disorders, sexual and gender identity disorders, physiological malfunction arising from mental factors, specific nonpsychotic mental disorders due to brain damage, intellectual disabilities. This is a heterogeneous category with regard to severity, and future work should use a more homogeneous grouping. Controlling for the situation where the patients visited any FFS psychiatrist in the previous two years prior to their index date, our result shows that patients suffering from anxiety disorder, substance use, schizophrenia and mood disorder have a higher chance of getting hospitalized comparing to those who suffer from "others" diseases. One possible interpretation for this result could be some of those disease types can not be treated properly without any acute care facilities. Whereas, if treated early, diseases like anxiety disorder or mood disorder can be managed by outpatient psychiatrists or by any community based mental health services.

Predictors interacting in the model, i.e, outpatient psychiatric visits and disease categories are depicted in Figure 4.10. The odds of having at least one hospitalization are consistently higher among the patients who did not receive outpatient psychiatrist care prior to the index date compared to those who did. However, differences in the odds of getting hospitalized are more evident for patients with schizophrenia, mood disorder and mental disorder due to substance use comparing to the patients suffering from "others" category of diseases . Narrower gaps between those with or without any prior psychiatric visits were observed for patients with behavioral disorders and anxiety disorder. Among patients without previous

outpatient psychiatric visits, and those having schizophrenia had the highest odds of having hospitalization. Schizophrenia patients who had any prior visits to psychiatrists are less likely to be hospitalized than those who did not visit any outpatient psychiatrist; however, there is no significant difference in the odds of getting hospitalized for the patients with anxiety, mental disorders due to substance use and behavioral disorders and have ever received outpatient psychiatrist care previously comparing to those who did not. These findings add to a growing literature on the inpatient hospitalization utilization research, highlighting the need for future investigation of interacting effect between outpatient psychiatric visit and disease types in relation to inpatient hospitalization rate for psychiatric conditions.

6. Conclusions

In this thesis, various counts models including Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, hurdle Poisson, hurdle negative binomial were considered for analyzing inpatient hospitalization data. We fit each of these models to the data based on large sets of linked administrative health data, in which the outcome variable was the count of repeated hospitalizations for psychiatric conditions. The negative binomial distribution better accounts for overdispersion in the outcome data compared to Poisson distribution. Hurdle and zero-inflated models account for excessive zero counts in the outcome variable. Examining the observed data, as well as fit statistics (AIC, Vuong's test as well as randomized quantile residuals), suggested that the distribution of the outcome variable was both overdispersed and zero-inflated. Models using the negative binomial distribution fit better than their corresponding models using the Poisson distribution, while zero-inflated and hurdle models fit better than their respective counterparts (Poisson, negative binomial). The fit statistics for the models (Tables 4.4-4.6) indicated that the hurdle negative binomial model provided the best fit. In the existing literature, comparison between zero-inflated and hurdle models remains understudied. In comparison to hurdle models, zero-inflated models consist of two sources of zero observations, "structural zeros" or non- at risk group that cannot score anything other than zero and "sampling zeros" that are part of the underlying sampling distribution (Poisson, or negative binomial). In the present study, all the study participants had at least one diagnosis of mental health condition and therefore could be at risk of being hospitalized. Taken together, our findings highlight the importance of accounting for both over-dispersion and zero-inflation, as well as considering and comparing both zero-inflated and hurdle models in modeling the count outcome data.

This thesis also leads to a better understanding of factors contributing to increased inpatient hospitalizations among patients with mental health conditions. The prevention of unnecessary hospitalizations has an impact on patients and caregivers–avoiding interruption in their lives and work activities, and saves cost for the health authority, as admissions are

the most expensive component of mental health budgets. Socio-demographic and clinical characteristics of the patients have been studied as possible influencing factors of readmission. Knowing the likely effect of these factors is useful for health professionals in order to detect high risk populations for prevention.

Our results are subject to some limitations. The index date in this study is the date of the first physician service claim on the Medical Services Plan (MSP) database between January 1, 2010 and December 31, 2011 reporting a psychiatric diagnosis (i.e., ICD-9 code 290-319). The codes that were used for this study to classify the diseases are based on the records of the index date, which can be somewhat different from the diagnosis code at the discharge. Moreover, while considering admissions over different study periods, we did not have the date of referral from a general physician and for which reason we could not estimate the waiting time.

The registered Indian status here was classified as ever declared as registered Indian or not. Registered Indians are Saskatchewan Health beneficiaries registered under Section 6 of the Indian Act and assigned a ten-digit number in the Indian Registry and have self-identified to the Saskatchewan Ministry of Health. Registered Indians are not eligible for Saskatchewan prescription drug benefits because they receive these benefits from the federal government for this reason self-declaration to the Ministry of Health as a registered Indian appears to be declining over time. Metis are not included in the registered Indian category. That missing information is one of the limitations of this study. The subject's residence is determined at the index date and reported using the categories listed above. There were no information whether those patients moved out from their place of residence at index date.

For this study, we could not consider some possible confounders like the socio-economic status of the patients, their income level and sources. The other possible confounders could be if the admitted patients were given psychiatric beds or not. As, admissions to any psychiatric bed could mean that patient was severely ill and had to stay longer time than usual. Again, the unavailability of psychiatric beds can lead to a premature discharge for some patients and increases the risk for a future readmission. Another possible confounder could be the information about the availability of mental health services at the community where the hospital is situated. Easy access to the community health services could prevent some

hospitalization for non-severe cases.

The "others" category consists of heterogenous group of diseases like: eating disorder, sextual preference disorder. Since this category was considered as the baseline category, comparing diseases with this heterogenous group could lead to incorrect conclusions. We used this categorization of disorders following a similar study among children by Rosychuk and colleagues [128].

Another limitation of our current study is that we could not consider the recurrent events for the hospitalizations. For recurrent events, there is intrinsic correlation between those events occurring in the same subject. The consequences of ignoring the recurrent nature of the data includes, the confidence intervals (CI) for the estimated rates could be artificially narrow and the null hypothesis is rejected more often than it should be [129]. To avoid this type of error, adjustments for within-individual correlation must be done. Our future work will include studying the within-subject correlations by applying the recurrent event data analysis appraoches.

A. Observed vs. Predicted Frequencies

Table A.1: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 3 months follow-up period.

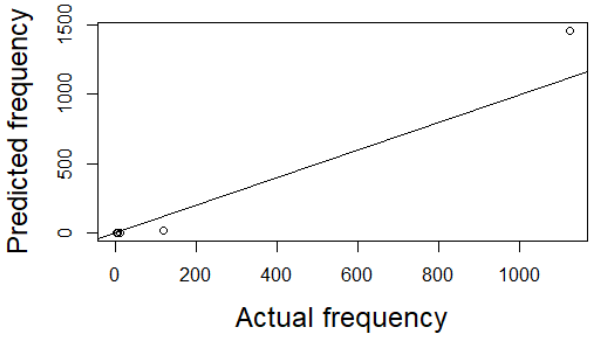
	Observed	Predicted frequencies					
Unique y	Frequencies	HNB	HP	ZINB	ZIP	NB	Poisson
0	199271	199271	199271	199279.3	199274.4	199272.2	199064
1	1124	1100.11	1077.46	1076.34	1068.52	1094.99	1460.30
2	118	131.06	159.92	143.05	164.14	126.14	12.52
3	10	25.28	23.92	28.01	25.08	27.36	0.20
4	1	6.41	3.88	7.02	4.04	8.72	0
5	2	1.96	0.67	2.08	0.67	3.54	0
6	4	0.68	0.11	0.70	0.11	1.68	0
7	2	0.26	0.01	0.25	0.01	0.88	0
8	2	0.11	0	0.10	0	0.05	0
9	2	0.04	0	0.04	0	0.03	0
10	1	0.02	0	0	0	0.01	0

Table A.2: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 6 months follow-up period.

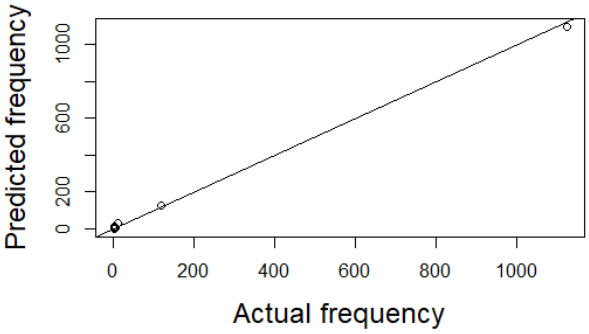
	Observed	Predicted frequencies					
Unique y	Frequencies	HNB	HP	ZINB	ZIP	NB	Poisson
0	198673	198673	198673	198688.8	198678.7	198674.8	198283
1	1597	1555.02	1508.17	1515.54	1497.78	1564.81	222.38
2	208	233.60	295.73	254.85	299.19	211.83	29.43
3	35	52.56	50.10	56.15	51.25	51.22	0.72
4	7	14.78	8.30	14.79	8.43	17.55	0.02
5	1	4.86	1.39	4.46	1.37	7.52	0
6	6	1.80	0.23	1.49	0.22	3.72	0
7	2	0.73	0.03	0.54	0.03	2.02	0
8	2	0.32	0	0.21	0	1.18	0
9	3	0.14	0	0.08	0	0.72	0
10	1	0.07	0	0.03	0	0.46	0
12	1	0.01	0	0	0	0.21	0
19	1	0	0	0	0	0	0

Table A.3: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 9 months follow-up period.

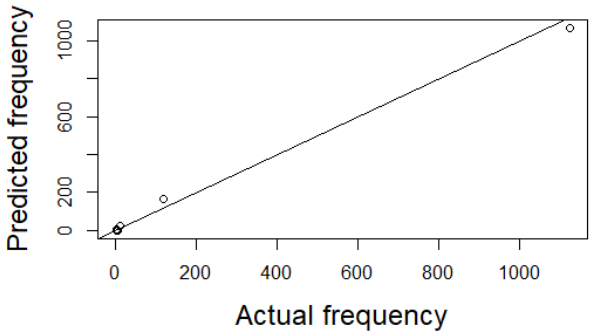
	Observed	Predicted frequencies					
Unique y	Frequencies	HNB	HP	ZINB	ZIP	NB	Poisson
0	198170	198170	198170	198191.62	198177.23	198172.45	197559.65
1	1913	1844.77	1764.15	1794.52	1754.48	1865.71	2861.24
2	282	326.73	427.84	352.28	430.40	294.70	47.63
3	60	84.41	86.92	89.66	87.41	78.46	1.41
4	20	26.73	16.97	26.86	16.66	28.58	0.06
5	4	9.78	3.33	9.12	3.12	12.80	0
6	5	4	0.64	3.42	0.57	6.58	0
7	4	1.78	0.12	1.39	0.10	3.71	0
8	2	0.85	0.02	0.60	0.02	2.24	0
9	4	0.43	0	0.27	0	1.42	0
11	1	0.12	0	0.06	0	0.64	0
12	1	0.07	0	0.03	0	0.44	0
15	1	0.01	0	0	0	0.17	0
16	1	0.01	0	0	0	0.13	0
19	1	0	0	0	0	0.06	0
21	1	0	0	0	0	0.04	0



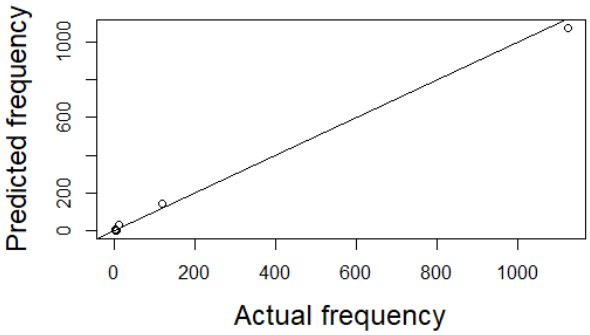
(a) Poisson



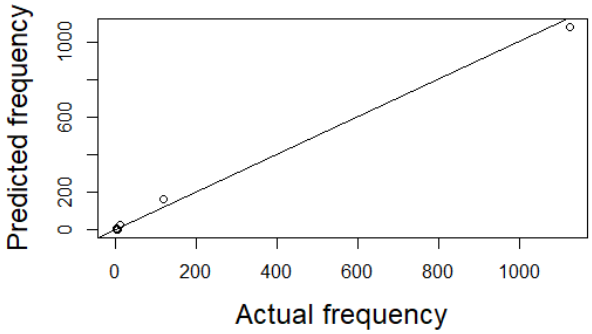
(b) NB



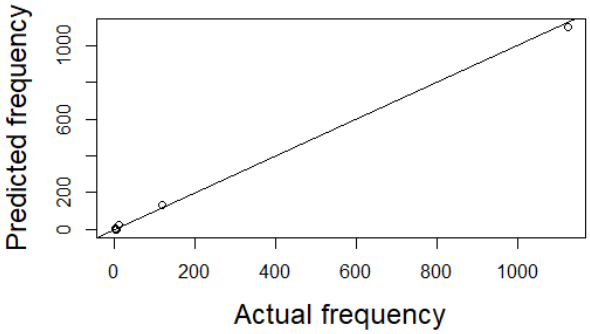
(c) ZIP



(d) ZINB

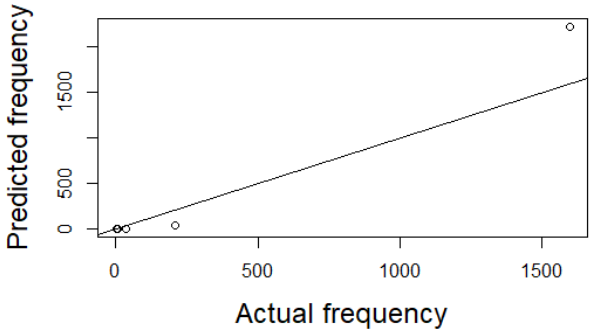


(e) HP

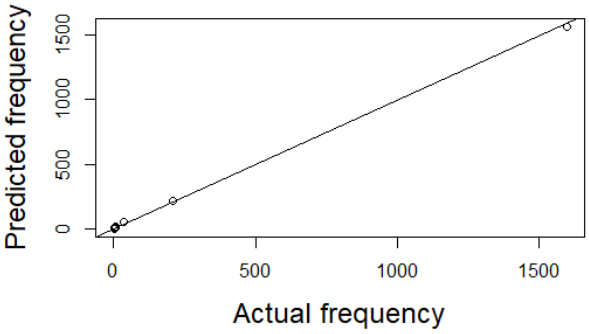


(f) HNB

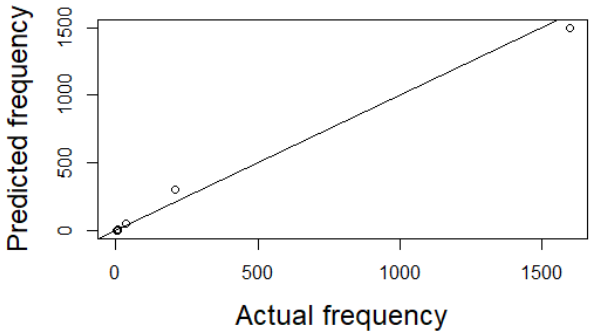
Figure A.1: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 3 months follow-up period.



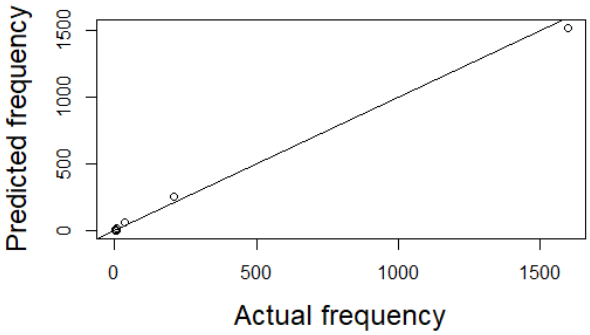
(a) Poisson



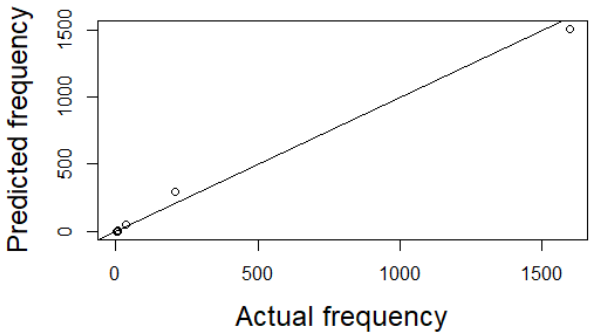
(b) NB



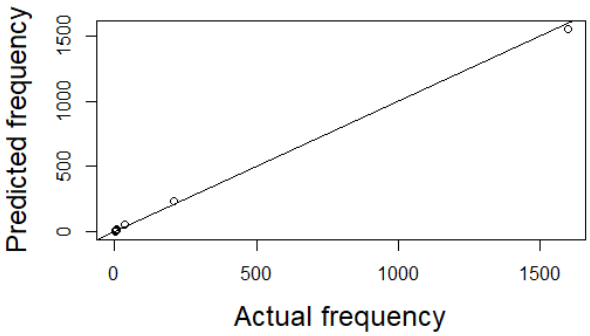
(c) ZIP



(d) ZINB

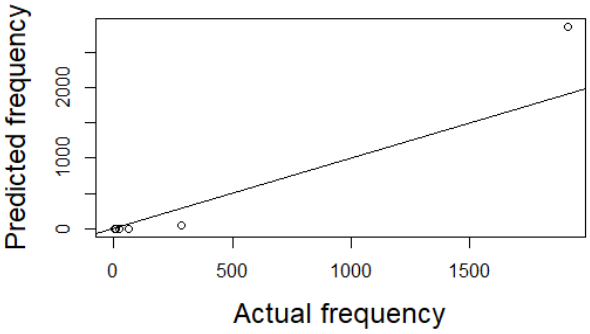


(e) HP

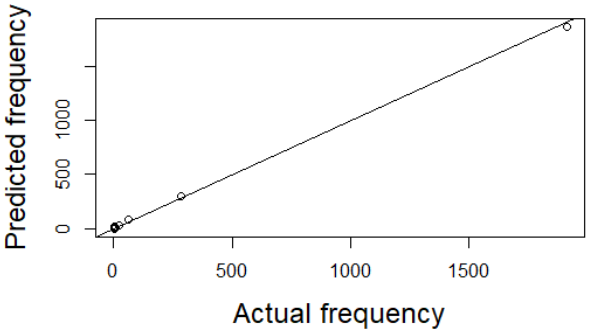


(f) HNB

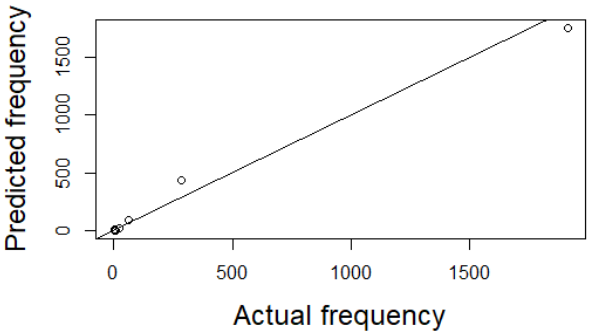
Figure A.2: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 6 months follow-up period.



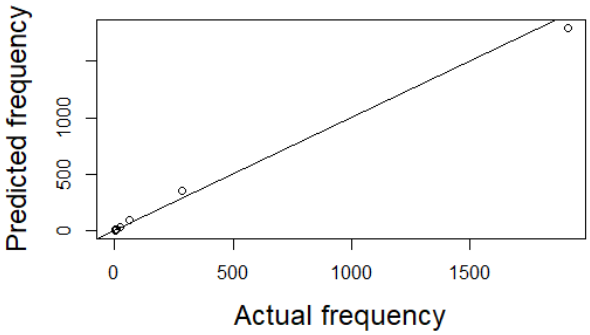
(a) Poisson



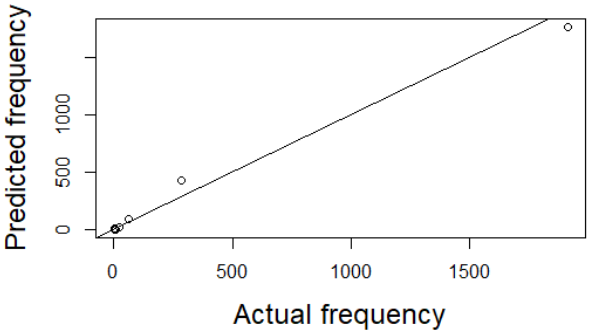
(b) NB



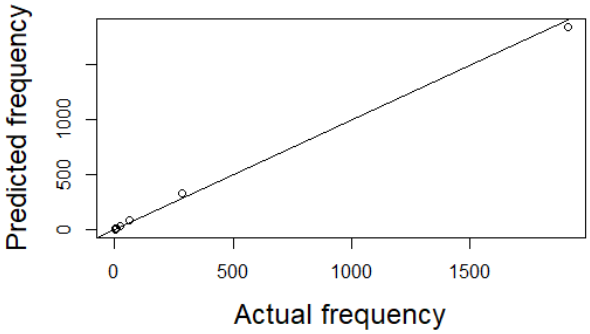
(c) ZIP



(d) ZINB



(e) HP



(f) HNB

Figure A.3: Observed vs. Predicted frequencies of the number of repeated inpatient hospitalizations for the 9 months follow-up period.

Bibliography

[1] World Health Organization. Mental health-a call for action by world health ministers, . URL https://www.who.int/mental_health/advocacy/en/Call_for_Action_MoH_Intro.pdf.

[2] Smetanin P., Stiff D., Briante C., Adair CE, Ahmad S., and Khan M. The life and economic impact of major mental illnesses in Canada: 2011 to 2041. Toronto: Riskanalytica, on behalf of the mental health commission of Canada; 2011.

[3] Nawaf M., Zhao H., and Fang L. J. Hospital readmissions for patients with mental illness in Canada. *Healthcare Quarterly*, 10(2):30–32, 2007.

[4] NHS Information Services Division Scotland ISD. Sparra mental disorder: Scottish patients at risk of readmission and admission (to psychiatric hospitals or units). a report on the development of sparra md and recommendations for its use, 2009. URL www.isdscotland.org/isd/servlet/FileBuffer?namedFileSPARRA-MD-report-final.Pdf&pContents.

[5] Mehta SA. Sixteen state pilot study: State mental health agency performance measures. indicator q11: Readmission to any state hospital within 30/180 days of discharge, 2001. URL www.mhsip.org/sixteenstate/0601Q11.pdf.

[6] Densen P. M., Shapiro S., and Einhorn M. Concerning high and low utilizers of service in a medical care plan, and the persistence of utilization levels over a three year period. *The Milbank Memorial Fund Quarterly*, 37(3):217–250, 1959.

[7] Taube C. A., Goldman H. H., Burus B. J., and Kessler L. G. High users of outpatient mental health services, i: Definition and characteristics. *The American journal of psychiatry*, 145(1):19, 1988.

[8] de Oliveira C., Cheng J., Vigod S., Rehm J., and Kurdyak P. Patients with high

mental health costs incur over 30 percent more costs than other high-cost patients. *Health Affairs*, 35(1):36–43, 2016.

[9] de Oliveira C., Cheng J., Rehm J., and Kurdyak P. The role of mental health and addiction among high-cost patients: a population-based study. *Journal of medical economics*, 21(4):348–355, 2018.

[10] Hensel J. M., Taylor V. H., Fung K., and Vigod S. N. Rates of mental illness and addiction among high-cost users of medical services in Ontario. *The Canadian Journal of Psychiatry*, 61(6):358–366, 2016.

[11] Cheng J., Rehm J., Kurdyak P., et al. The economic burden of chronic psychotic disorders in Ontario. *The journal of mental health policy and economics*, 19(4):181–192, 2016.

[12] Fortney J. C., Xu S., and Dong F. Community-level correlates of hospitalizations for persons with schizophrenia. *Psychiatric Services*, 60(6):772–778, 2009.

[13] Machado V., Leonidas C., Santos MA., and Souza J. Psychiatric readmission: an integrative review of the literature. *International Nursing Review*, 59:44757, 2012.

[14] Durbin J., Lin E., Layne C., and Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? *J Behav Health Serv Res*, 34:137–50, 2007.

[15] Rumball-Smith J. and Hider P. The validity of readmission rate as a marker of the quality of hospital care, and a recommendation for its definition. *NZ Med J*, 122:63–70, 2009.

[16] AHRQ. Health plan and psychiatric hospitals reduce readmissions by reviewing data and developing strategies to improve postdischarge care, 2009. URL https://innovations.ahrq.gov/profiles/health-plan-and-psychiatric-hospitals-reduce-readmissionsreviewing-data-and-develop

[17] Bradley EH., Curry L., Horwitz LI., Sipsma H., Thompson JW., and Elma M. Contemporary evidence about hospital strategies for reducing 30-day readmissions: a national study. *J Am Coll Cardiol*, 60:60714, 2012.

[18] Anderson M., Revie C. W., Quail J. M., Wodchis W., de Oliveira C., Osman M., Baetz M., McClure J., Stryhn H., Buckeridge D., et al. The effect of socio-demographic factors on mental health and addiction high-cost use: a retrospective, population-based study in Saskatchewan. *Canadian Journal of Public Health*, 109(5-6):810–820, 2018.

[19] Jorid Kalseth, Eva Lassemo, Kristian Wahlbeck, Peija Haaramo, and Jon Magnussen. Psychiatric readmissions and their association with environmental and health system characteristics: a systematic review of the literature. *BMC psychiatry*, 16(1):376, 2016.

[20] Appleby L., Luchins D. J., Freels S., Smith M. E., and Wasmer D. The impact of immigration on psychiatric hospitalization in Illinois from 1993 to 2003. *Psychiatric Services*, 59(6):648–654, 2008.

[21] Klinkenberg W. Dean and Calsyn R. J. Predictors of receipt of aftercare and recidivism among persons with severe mental illness: a review. *Psychiatric Services*, 1996.

[22] Lin CH., Chen MC., Chou LS., Lin CH., Chen CC., and Lane HY. Time to rehospitalization in patients with major depression vs. those with schizophrenia or bipolar i disorder in a public psychiatric hospital. *Psychiatry research*, 180(2-3):74–79, 2010.

[23] Weiden P., Aquila R., and Standard J. Atypical antipsychotic drugs and long-term outcome of schizophrenia. *The Journal of Clinical Psychiatry*, 1996.

[24] Chang C. K., Hayes R. D., Perera G., Broadbent M., TM., Fernandes A. C., W. E. Lee, Hotopf M., and Stewart R. Life expectancy at birth for people with serious mental illness and other major disorders from a secondary mental health care case register in London. *PloS one*, 6(5):e19590, 2011.

[25] Vos T., Flaxman A. D., Naghavi M., Lozano R., Michaud C., Ezzati M., Shibuya K., J. A. Salomon, Abdalla S., Aboyans V., et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, 380(9859):2163–2196, 2012.

[26] Murray C. JL., Vos T., Lozano R., Naghavi M., Flaxman A. D., Michaud C., Ezzati M., Shibuya K., Salomon J. A., Abdalla S., et al. Disability-adjusted life years (DALYs)

for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet*, 380(9859):2197–2223, 2012.

[27] Li Y., Glance L. G., Cai X., and Mukamel D. B. Mental illness and hospitalization for ambulatory care sensitive medical conditions. *Medical care*, pages 1249–1256, 2008.

[28] Bouza C., López-Cuadrado T., and Amate J. María. Hospital admissions due to physical disease in people with schizophrenia: a national population-based study. *General hospital psychiatry*, 32(2):156–163, 2010.

[29] Marder S. R., Essock S. M., Miller A. L., Buchanan R. W., Casey D. E., Davis J. M., Kane J. M., Lieberman J. A., Schooler N. R., Covell N., et al. Physical health monitoring of patients with schizophrenia. *American Journal of Psychiatry*, 161(8):1334–1349, 2004.

[30] McManus S., Meltzer H., and Campion J. Cigarette smoking and mental health in England: Data from the adult psychiatric morbidity survey 2007. *London, UK: National Centre for Social Research*, 2010.

[31] Addington D. E., Beck C., Wang J., Adams B., Pryce C., Zhu H., Kang J., and McKenzie E. Predictors of admission in first-episode psychosis: developing a risk adjustment model for service comparisons. *Psychiatric services*, 61(5):483–488, 2010.

[32] Gorwood P. Factors associated with hospitalisation of patients with schizophrenia in four European countries. *European Psychiatry*, 26(4):224–230, 2011.

[33] Hoblyn J. C., Balt S. L., Woodard S. A., and Brooks III J. O. Substance use disorders as risk factors for psychiatric hospitalization in bipolar disorder. *Psychiatric Services*, 60(1):50–55, 2009.

[34] Schubart CD., Boks MPM., Breetvelt EJ., Van Gastel WA., Groenwold RHH., Ophoff RA., Sommer I. EC., and Kahn RS. Association between cannabis and psychiatric hospitalization. *Acta Psychiatrica Scandinavica*, 123(5):368–375, 2011.

[35] Sipos A., Harrison G., Gunnell D., Amin S., and Singh S. P. Patterns and predictors of hospitalisation in first-episode psychosis: prospective cohort study. *The British Journal of Psychiatry*, 178(6):518–523, 2001.

[36] John L. H., Offord D. R., Boyle M. H., and Racine Y. A. Factors predicting use of mental health and social services by children 6–16 years old: Findings from the Ontario child health study. *American Journal of Orthopsychiatry*, 65(1):76–86, 1995.

[37] Palin J. L., Goldner E. M., Koehoorn M., and Hertzman C. Primary mental health care visits in self-reported data versus provincial administrative records. *Health reports*, 22(2):41, 2011.

[38] Offord D. R., Boyle M. H., Fleming J. E., Blum H. M., and Rae Grant N. I. Summary of selected results. *The Canadian Journal of Psychiatry*, 34(6):483–491, 1989.

[39] Waddell C., Offord D. R., Shepherd C., A., Hua J. M., and McEwan K. Child psychiatric epidemiology and Canadian public policy-making: the state of the science and the art of the possible. *The Canadian Journal of Psychiatry*, 47(9):825–832, 2002.

[40] Mason M. A. and Gibbs J. T. Patterns of adolescent psychiatric hospitalization: Implications for social policy. *American Journal of Orthopsychiatry*, 62(3):447–457, 1992.

[41] Fontanella C. A. The influence of clinical, treatment, and healthcare system characteristics on psychiatric readmission of adolescents. *American Journal of Orthopsychiatry*, 78(2):187–198, 2008.

[42] James S., Charlemagne S. J., Gilman A. B., Alemi Q., Smith R. L., Tharayil P. R., and Freeman K. Post-discharge services and psychiatric rehospitalization among children and youth. *Administration and Policy in Mental Health and Mental Health Services Research*, 37(5):433–445, 2010.

[43] Romansky J. B., Lyons J. S., Lehner R. K., and West C. M. Factors related to psychiatric hospital readmission among children and adolescents in state custody. *Psychiatric Services*, 54(3):356–362, 2003.

[44] Luber M. P., Meyers B. S., Williams-Russo P. G., Hollenberg J. P., DiDomenico T. N., Charlson M. E., and Alexopoulos G. S. Depression and service utilization in elderly primary care patients. *The American Journal of Geriatric Psychiatry*, 9(2):169–176, 2001.

[45] Beekman A. TF, Penninx B. WJH, Deeg D. JH, Beurs E. de, Geerlings S. W., and Tilburg W. van. The impact of depression on the well-being, disability and use of services in older adults: a longitudinal perspective. *Acta Psychiatrica Scandinavica*, 105(1):20–27, 2002.

[46] Fischer L. R., Wei F., Rolnick S. J., Jackson J. M., Rush W. A., Garrard J. M., Nitz N. M., and Luepke L. J. Geriatric depression, antidepressant treatment, and healthcare utilization in a health maintenance organization. *Journal of the American Geriatrics Society*, 2002.

[47] Anttila S. Depressive symptoms and general hospital care in the elderly a population-based study. *Nordisk Psykiatrisk Tidsskrift*, 45(5):345–349, 1991.

[48] Canadian Institute for Health Information. Hospital mental health services in Canada, 2003-2004, 2006. URL https://secure.cihi.ca/free_products/HMHDB_2003-2004_e.pdf.

[49] Carrière G. M., Bougie E., Kohen D., Rotermann M., and Sanmartin C. *Acute care hospitalization by Aboriginal identity, Canada, 2006 through 2008*. Statistics Canada, 2016.

[50] Pong R. W., DesMeules M., Heng D., Lagacé C., Guernsey J. R., Kazanjian A., Manuel D., Pitblado JR, Bollman R., Koren I., et al. Patterns of health services utilization in rural Canada. *Chronic diseases and injuries in Canada*, 31:1–36, 2011.

[51] World Health Organization. Gender and women's mental health. https://www.who.int/mental_health/prevention/genderwomen/en/, .

[52] Rüesch P., Meyer PC, and Hell D. Who is rehospitalized in a psychiatric hospital? psychiatric hospitalization rates and social indicators in the Zurich canton (Switzer-

land). *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany))*, 62(3):166–171, 2000.

[53] Chuan-Hsiung L., Wen-Ling C., Chih-Ming L., Lee RN, Ming-Chung K., Chung-Yi L., et al. Predictors of psychiatric readmissions in the short-and long-term: a population-based study in Taiwan. *Clinics*, 65(5):481–489, 2010.

[54] Priebe S., Katsakou C., Amos T., Leese M., Morriss R., and Rose D. Patients views and readmissions 1 year after involuntary hospitalization. *Br J Psychiatry*, 194:4954, 2009.

[55] Brennan PL., Kagay CR., Geppert JJ., and Moos RH. Elderly Medicare inpatients with substance use disorders: characteristics and predictors of hospital readmissions over a four-year interval. *J Stud Alcohol*, 61:8915, 2000.

[56] Monnelly EP. Instability before discharge and previous psychiatric admissions as predictors of early readmission. *Psychiat Serv*, 48:15846, 1997.

[57] Ono T., Tamai A., Takeuchi D., and Tamai Y. Factors related to readmission to a ward for dementia patients: Sex differences. *PsychiatClin Neuros*, 65:4908, 2011.

[58] Craig TJ., Fennig S., Tanenberg-Karant M., and Bromet EJ. . Rapid versus delayed readmission in first-admission psychosis: quality indicators for managed care? . *Ann Clin Res*, 12:2338, 2000.

[59] Morrow-Howell NL., Proctor EK., Blinne WR., Rubin EH., Saunders JA., and Rozario PA. Post-acute dispositions of older adults hospitalized for depression. *Aging Ment Health*, 10:35261, 2006.

[60] Sanguineti V. R., Samuel S. E., Schwartz S. L., and Robeson M. R. Retrospective study of 2,200 involuntary psychiatric admissions and readmissions. *The American Journal of Psychiatry*, 153(3):392, 1996.

[61] Sullivan G., Young AS., and Morgenstern H. Behaviors as risk factors for rehospitalization: implications for predicting and preventing admissions among the seriously mentally ill. *Social Psychiatry and Psychiatric Epidemiology*, 32(4):185–190, 1997.

[62] Vogel S. and Huguelet P. Factors associated with multiple admissions to a public psychiatric hospital. *Acta Psychiatrica Scandinavica*, 95(3):244–253, 1997.

[63] Havassy B. E. and Hopkin J. T. Factors predicting utilization of acute psychiatric inpatient services by frequently hospitalized patients. *Psychiatric Services*, 40(8):820–823, 1989.

[64] Haywood T. W., Kravitz H. M., Grossman L. S., Cavanaugh Jr J. L., et al. Predicting the" revolving door" phenomenon among patients with schizophrenic, schizoaffective, and affective disorders. *The American journal of psychiatry*, 152(6):856, 1995.

[65] Gardner W., Mulvey E. P., and Shaw E. C. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin*, 118(3):392, 1995.

[66] Lyon M., Lyon N., and Magnusson M. S. The importance of temporal structure in analyzing schizophrenic behavior: some theoretical and diagnostic implications. *Schizophrenia Research*, 13(1):45–56, 1994.

[67] Morgan V., Castle D., Page A., Fazio S., Gurrin L., Burton P., Montgomery P., and Jablensky A. Influenza epidemics and incidence of schizophrenia, affective disorders and mental retardation in western Australia: no evidence of a major effect. *Schizophrenia Research*, 26(1):25–39, 1997.

[68] Selten J. P., Slaets J., and Kahn R. S. Schizophrenia in surinamese and dutch antillean immigrants to the Netherlands: evidence of an increased incidence. *Psychological Medicine*, 27(4):807–811, 1997.

[69] Tansella M., Micciolo R., Biggeri A., Bisoffi G., and Balestrieri M. Episodes of care for first-ever psychiatric patients a long-term case-register evaluation in a mainly urban area. *The British Journal of Psychiatry*, 167(2):220–227, 1995.

[70] Evans M. N. hastings, and b. peacock, statistical distributions, 1993.

[71] Herrendörfer G. Johnson, nl, kotz, s.; kemp, aw: Univariate discrete distributions. *Biometrical Journal*, 36(2):164–164, 1994.

[72] Selvin S. *Practical biostatistical methods*, volume 1. Brooks/Cole, 1995.

[73] McKenzie D. P., Mackinnon A. J., Martindale C., and Clarke D. M. Modelling the frequency of psychiatric admissions. *International Journal of Methods in Psychiatric Research*, 7(3):136–41, 1998.

[74] Eaton Jr W. W. Mental hospitalization as a reinforcement process. *American sociological review*, pages 252–260, 1974.

[75] Smeeton N. C. Distribution of episodes of mental illness in general practice: results from the second national morbidity survey. *Journal of Epidemiology Community Health*, 40 (2):130–133, 1986.

[76] Greenwood M. and Yule GU. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal statistical society*, 83(2):255–279, 1920.

[77] Ruef M. and Scott W. R. A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments. *Administrative science quarterly*, pages 877–904, 1998.

[78] Modestin J. and Würmle O. Role of modelling in in-patient suicide: a lack of supporting evidence. *The British Journal of Psychiatry*, 155(4):511–514, 1989.

[79] Glynn R. J., Stukel T. A., Sharp S. M., Bubolz T. A., Freeman J. L., and Fisher E. S. Estimating the variance of standardized rates of recurrent events, with application to hospitalizations among the elderly in new England. *American Journal of Epidemiology*, 137(7):776–786, 1993. doi: 10.1093/oxfordjournals.aje.a116738. URL http://dx.doi.org/10.1093/oxfordjournals.aje.a116738.

[80] Gardner MJ MJ. and Altman DG. *Statistics with Confidence: Confidence Intervals and Statistical Guidelines*. London: British Medical Journal, 1989.

[81] Van den Broek J. A score test for zero inflation in a Poisson distribution. *Biometrics*, pages 738–743, 1995.

[82] Mullahy J. Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365, 1986.

[83] Zurr AF., Ieno EN., Walker NJ., Saveliev AA., and Smith GM. Zero-truncated and zero-inflated models for count data. *Mixed effects models and extensions in ecology with R. Springer, New York, New York, USA*, pages 261–293, 2009.

[84] Lyons J. S., O'mahoney M. T., Miller S. I., Neme J., Kabat J., and Miller F. Predicting readmission to the psychiatric hospital in a managed care environment: implications for quality indicators. *American Journal of Psychiatry*, 154(3):337–340, 1997.

[85] Durbin J., Lin E., Layne C., and Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? *The Journal of Behavioral Health Services & Research*, 34(2):137–150, 2007.

[86] Akaike H. Theory and an extension of the maximum likelihood principal. In *International symposium on information theory. Budapest, Hungary: Akademiai Kaiado*, 1973.

[87] Vuong Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

[88] Dunn P. K. and Smyth G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244, 1996.

[89] Johnson N. L., KOTZ S., and KEMP A. W. Univariate discrete distributions. second edition. *Biometrical Journal*, 36(2):164–164, 1992. doi: 10.1002/bimj.4710360207. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710360207.

[90] Hilbe J. M. *Negative binomial regression*. Cambridge University Press, 2011.

[91] Mwalili S. M., Lesaffre E., and Declerck D. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical methods in medical research*, 17(2):123–39, 2008. doi: 10.1177/0962280206071840. URL https://app.dimensions.ai/details/publication/pub.1063903935. Exported from https://app.dimensions.ai on 2018/12/04.

[92] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. 34(1):1, 1992. doi: 10.2307/1269547. URL https://app.dimensions.ai/details/publication/pub.1069421209. Exported from https://app.dimensions.ai on 2018/12/04.

[93] Cameron A. C. and Trivedi P. K. *Regression analysis of count data*, volume 53. Cambridge university press, 2013.

[94] Zuur A. F., Ieno E. N., Walker N. J., Saveliev A. A., and Smith G. M. *Zero-Truncated and Zero-Inflated Models for Count Data*, pages 261–293. Springer New York, New York, NY, 2009. ISBN 978-0-387-87458-6. doi: 10.1007/978-0-387-87458-6_11. URL https://doi.org/10.1007/978-0-387-87458-6_11.

[95] A. Agresti. *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley, 2012. ISBN 9780470463635. URL https://books.google.ca/books?id=UOrr47-2oisC.

[96] Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482, 1943.

[97] Swets J. A. Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293, 1988.

[98] Bradley A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[99] Detector performance analysis using roc curves. https://www.mathworks.com/help/phased/examples/detector-performance-analysis-using-roc-curves.html.

[100] Øiesvold T., Saarento O., Sytema S., Vinding H., Göstas G., Lönnerberg O., Muus S., Sandlung M., and Hansson L. Predictors for readmission risk of new patients: the Nordic comparative study on sectorized psychiatry. *Acta psychiatrica scandinavica*, 101 (5):367–373, 2000.

[101] Silva N. C., Bassani D. G., and Palazzo L. S. A case-control study of factors associated with multiple psychiatric readmissions. *Psychiatric Services*, 60(6):786–791, 2009.

[102] Aboriginal Healing Foundation. From Truth to Reconciliation: Transforming the Legacy of Residential Schools. Ottawa, ON: Aboriginal Healing Foundation, 2002. URL http://www.ahf.ca/downloads/truth-to-reconciliation.pdf.

[103] Adelson N. The embodiment of inequity. *Canadian Journal of Public Health*, 96(2): S45–S61, 2005.

[104] Bombay A., Matheson K., and Anisman H. Intergenerational trauma: Convergence of multiple processes among first nations peoples in Canada. *International Journal of Indigenous Health*, 5(3):6–47, 2009.

[105] Bombay A. M., Matheson K., and Anisman H. The intergenerational effects of Indian residential schools: Implications for the concept of historical trauma. *Transcultural psychiatry*, 51(3):320–338, 2014.

[106] Truth and Reconciliation Commission of Canada. *Honouring the truth, reconciling for the future: Summary of the final report of the Truth and Reconciliation Commission of Canada*. Truth and Reconciliation Commission of Canada, 2015.

[107] Caron J., Latimer É., and Tousignant M. Predictors of psychological distress in low-income populations of Montreal. *Canadian Journal of Public Health*, 98(1):S35–S44, 2007.

[108] Firestone M., Smylie J., Maracle S., McKnight C., Spiller M., and OCampo P. Mental health and substance use in an urban first nations population in Hamilton, Ontario. *Canadian Journal of Public Health*, 106(6):e375–e381, 2015.

[109] Allan B. and Smylie J. *First Peoples, Second Class Treatment: The Role of Racism in the Health and Well-being of Indigenous Peoples in Canada, Discussion Paper*. Wellesley Institute, 2015.

[110] Statistics Canada. 2006 census of Canada: Special interest profiles. URL https://www12.statcan.gc.ca/census-recensement/2006/dp-pd/prof/sip/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=97446&PRID=0&PTYPE=97154&S=0&SHOWALL=0&SUB=0&Temporal=2006&THEME=73&VID=0&VNAMEE=&VNAMEF=.

[111] Bingham B. Aboriginal community-based primary health care research: Developing community-driven primary health care research priorities. *Surrey, British Columbia: Aboriginal Health Services, Fraser Health*, 2013.

[112] Browne A. J., Smye V. L., Rodney P., Tang S. Y., Mussell B., and O'Neil J. Access to primary care from the perspective of aboriginal patients at an urban emergency department. *Qualitative Health Research*, 21(3):333–348, 2011.

[113] Merikangas K. R., Nakamura E. F., and Kessler R. C. Epidemiology of mental disorders in children and adolescents. *Dialogues in clinical neuroscience*, 11(1):7, 2009.

[114] Costello E. J., Mustillo S., Erkanli A., Keeler G., and Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of general psychiatry*, 60(8):837–844, 2003.

[115] Costello E. J., Egger H., and Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(10):972–986, 2005.

[116] Brauner C. B. and Stephens C. B. Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations. *Public health reports*, 121(3):303–310, 2006.

[117] McLean C. P., Asnaani A., Litz B. T., and Hofmann S. G. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *Journal of psychiatric research*, 45(8):1027–1035, 2011.

[118] Hawton K. Sex and suicide: Gender differences in suicidal behaviour. *The British Journal of Psychiatry*, 177(6):484–485, 2000.

[119] World Health Organization. Management of Substance Abuse Unit. *Global status report on alcohol and health, 2014*. World Health Organization, 2014.

[120] Maguen S., Cohen B., Cohen G., Madden E., Bertenthal D., and Seal K. Gender differences in health service utilization among Iraq and Afghanistan veterans with post-traumatic stress disorder. *Journal of Women's Health*, 21(6):666–673, 2012.

[121] Prince J. D., Akincigil A., Kalay E., Walkup J. T., Hoover D. R., Lucas J., Bowblis J., and Crystal S. Psychiatric rehospitalization among elderly persons in the United States. *Psychiatric Services*, 59(9):1038–1045, 2008.

[122] Husted J. and Jorgens A. Best practices: population density as a factor in the rehospitalization of persons with serious and persistent mental illness. *Psychiatric Services*, 51(5):603–605, 2000.

[123] Stahler G. J., Mennis J., Cotlar R., and Baron D. A. The influence of neighborhood environment on treatment continuity and rehospitalization in dually diagnosed patients discharged from acute inpatient care. *American Journal of Psychiatry*, 166(11):1258–1268, 2009.

[124] Zhang J., Harvey C., and Andrew C. Factors associated with length of stay and the risk of readmission in an acute psychiatric inpatient facility: a retrospective study. *Australian and New Zealand Journal of Psychiatry*, 45(7):578–585, 2011.

[125] Beecham J. Access to mental health supports in England: crisis resolution teams and day services. *International Journal of Law and Psychiatry*, 28(5):574–587, 2005.

[126] Gaddini A., Ascoli M., and Biscaglia L. Mental health care in Rome. *European psychiatry*, 20:S294–S297, 2005.

[127] Guo S., Biegel D. E., Johnsen J. A., and Dyches H. Assessing the impact of community-based mobile crisis services on preventing hospitalization. *Psychiatric Services*, 52(2): 223–228, 2001.

[128] Newton A. S., Rathee S., Grewal S., Dow N., and Rosychuk R. J. Childrens mental health visits to the emergency department: factors affecting wait times and length of stay. *Emergency medicine international*, 2014, 2014.

[129] Amorim L. DAF and Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1):324–333, 12 2014. ISSN 0300-5771. doi: 10.1093/ije/dyu222. URL https://doi.org/10.1093/ije/dyu222.