Evaluating the Sociocultural Risks for Suicide Attempt in Schizophrenia

by

Nuan Chamera Hettige

A thesis submitted in conformity with the requirements for the degree of Master of Science
Institute of Medical Science
University of Toronto

© Copyright by Nuwan Chamera Hettige 2016
Evaluating the Sociocultural Risks for Suicide Attempt in Schizophrenia

Nuwan Chamera Hettige
Master of Science
Institute of Medical Science
University of Toronto
2016

Abstract

Sociocultural factors involved in immigration and ethnicity are believed to independently influence the risk for schizophrenia and suicidal behaviour. Schizophrenia patients are known to be at an increased risk for suicide. Though many risk factors have been identified, few have any significant clinical impact on predicting suicide. Whether immigration and ethnicity are potential risk factors for suicide attempt in schizophrenia has yet to be investigated. Schizophrenia patients with clear suicide attempt history, immigration history, self-reported ethnicity, and other sociocultural and clinical variables were recruited to test whether independently, or synthesized in a classification algorithm, these variables can accurately predict a history of suicide attempt. Both immigration and ethnicity had a non-significant association with suicide attempt. Inclusion of these predictor variables also did little to improve classification accuracy in our algorithms. Taken together, we found no evidence that these sociocultural factors have a significant influence on the risk for suicide attempt in schizophrenia.
Acknowledgments

First and foremost, I would like to sincerely thank my supervisor Dr. Vincenzo De Luca for giving me the chance to begin my research career in his lab while still an undergraduate student. I consider myself very fortunate to have been given the chance to work in your lab. The commitment you have to mentoring your students is unparalleled and I am grateful to have been given the opportunity to learn all that I could from you.

To the dedicated members of my Progress Advisory Committee: Dr. Arun Ravindran, Dr. Isabelle Boileau, and Dr. Isaac Sakinofsky; thank you for your positive encouragement and kindness. Your mentorship and scientific support have helped me improve every step of the way.

Furthermore, I could not have come this far without the support and friendship of my fellow lab member, Ali Bani-Fatemi, who shared the same goal of getting through graduate school unscathed. Thank you for your wisdom and friendship over the last four years. Because of you I was able to navigate the roadblocks of my degree with greater ease.

Being a long-standing member of the lab, I have seen many students and volunteers come and go; all of whom have left a lasting impression on me and helped make my job every bit easier. A special thank you to Gina Polsinelli, Aaron Howe, Roy Kamhi, Michelle Matmari, Tiffany Leung, Jiali Song, Thanara Rajakulendran, Jermeen Baddour, Gustavo de Moraes, Alex Pereira, and Arthur Koga for all your help and contributions.

During my time as a graduate student, I was fortunate enough to have made many friends along the way who always made it an enjoyable experience to be at the lab. To the Originals: Eric
Huang, Amanda Lisoway, Venuja Sriretnakumar, and Victoria Marshe - thank you for all the good times throughout grad school. It was never a dull moment with all of you around.

To my best friend, Manal Al-Ayad, without your unconditional love and support I would not have become the person I am today and made it this far. Thank you for always supporting me and believing in me no matter what. You are my everything. I hope I continue to make you proud as we continue into our future together.

Lastly, all my work and achievements leading up to this point are dedicated to my loving family. To my amazing parents - my mother who always made sure I would never spend a day hungry and my late father, Ariyadasa Hettige, who taught me what it meant to be a responsible, kind, and hardworking human being. This is for you.
Contributions

Nuwan Hettige is the sole author of this thesis. All components of this thesis were completed in whole or in part by the author. The following contributions by other individuals are formally and inclusively acknowledged:

Dr. Vincenzo De Luca (Graduate Supervisor) – mentorship; laboratory resources; assistance in study design and planning; critical feedback on writing; clinical assessment training.

Dr. Isaac Sakinofsky (Program Advisory Committee Member) – guidance and assistance in navigating the scientific literature; editing and revisions on many aspects of this thesis.

Dr. Arun Ravindran (Program Advisory Committee Member) – mentorship; guidance in interpretation of results and thesis preparation.

Dr. Isabelle Boileau (Program Advisory Committee Member) – mentorship; guidance in interpretation of results and thesis preparation.

Dr. James L. Kennedy (Collaborator) – provided participants with clinical and genetic data which comprised a portion of the sample in chapter 3.

Gina Polsinelli, Aaron Howe, Roy Kamhi, Michelle Matmari, Jiali Song, Tiffany Leung (Research Assistants) - assistance in patient recruitment, assessments, data entry, chart review, consenting participants, and completing clinical assessments.

Nikhil Bhagwat - assistance in developing the Python code for statistical analyses in Chapter 4.

My work was supported by the Margaret & Howard GAMBLE Research grant and Open Fellowship Award both offered by the University of Toronto, and a Canadian Institutes of Health Research (CIHR) Travel Award to present portions of this thesis.
# Table of Contents

Abstract ............................................................................................................................... ii  
Acknowledgments .............................................................................................................. iii  
Contributions ...................................................................................................................... v  
Table of Contents ................................................................................................................ vi  
List of Abbreviations ......................................................................................................... ix  
List of Tables ....................................................................................................................... x  
List of Figures ...................................................................................................................... xi  
List of Appendices ............................................................................................................. xiii  
Chapter 1 ............................................................................................................................. 1  

## 1 Literature Review ........................................................................................................... 1  

### 1.1 Schizophrenia ............................................................................................................. 1  

#### 1.1.1 Overview and Clinical Features ............................................................................. 1  

#### 1.1.2 Epidemiology ........................................................................................................ 3  

#### 1.1.3 Prognosis and Clinical Management .................................................................... 4  

### 1.2 Impact of Immigration on the Development of Schizophrenia ............................... 6  

#### 1.2.1 Incidence of Psychosis among Immigrants ......................................................... 6  

#### 1.2.2 Difficulty in Acculturation .................................................................................. 7  

#### 1.2.3 Cultural and Ethnic Communities ...................................................................... 9  

#### 1.2.4 Low Socioeconomic Status ............................................................................... 10  

#### 1.2.5 Selective Immigration ....................................................................................... 11  

#### 1.2.6 Population Genetics ............................................................................................ 11  

### 1.3 Suicidal Behaviour .................................................................................................... 12  

#### 1.3.1 Epidemiology ...................................................................................................... 12  

#### 1.3.2 Immigration and Suicide .................................................................................... 14  

#### 1.3.3 Suicide Terminology and Definitions ................................................................ 16  

#### 1.3.4 Investigating Suicide Attempts ........................................................................... 17  

### 1.4 Suicide in Schizophrenia ............................................................................................ 18  

#### 1.4.1 Prevalence and Incidence Rates ......................................................................... 18  

#### 1.4.2 Risk Factors for Suicide in Schizophrenia ......................................................... 19  

### 1.4.2.1 Clinical Demographics ..................................................................................... 20
Chapter 4 Classification of Suicide Attempt History in Schizophrenia using Sociocultural and Clinical Features ................................................................. 70

4.1 Abstract ............................................................................................................ 70
4.2 Introduction ....................................................................................................... 71
4.3 Materials and Methods .................................................................................. 73
   4.3.1 Study Sample ............................................................................................. 73
   4.3.2 Clinical Assessments ................................................................................ 74
   4.3.3 Predictor Variable Selection .................................................................... 74
   4.3.4 Classification Algorithms ....................................................................... 75
   4.3.5 Model Testing and Cross-Validation ....................................................... 76
4.4 Results .............................................................................................................. 79
4.5 Discussion ........................................................................................................ 87
4.6 Conclusion ....................................................................................................... 90

Chapter 5 General Discussion ............................................................................... 92

5.1 Discussion ....................................................................................................... 92
   5.1.1 Immigration and Ethnicity as Predictors .............................................. 94
   5.1.2 Suicide Classification ............................................................................ 97
5.2 Limitations ....................................................................................................... 100
5.3 Conclusions .................................................................................................... 105
5.4 Future Directions ........................................................................................... 107
   5.4.1 Longitudinal Assessment of Immigration ........................................... 107
   5.4.2 Comparison of Clinical Judgment and Machine learning .................. 109
   5.4.3 Improved Phenotype for Classification ................................................. 111

References ............................................................................................................. 113

Appendices ............................................................................................................ 137
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-HIAA</td>
<td>5-hydroxyindoleacetic acid</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>CAMH</td>
<td>Centre for Addiction and Mental Health</td>
</tr>
<tr>
<td>CEU</td>
<td>North/Western Europeans from Utah, USA</td>
</tr>
<tr>
<td>CHB</td>
<td>Han Chinese East Asians</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal Fluid</td>
</tr>
<tr>
<td>CTQ</td>
<td>Childhood Trauma Questionnaire</td>
</tr>
<tr>
<td>DSM-IV</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, 4th Edition</td>
</tr>
<tr>
<td>DSM-V</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, 5th Edition</td>
</tr>
<tr>
<td>FEP</td>
<td>First-Episode Psychosis</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-Wide Association Study</td>
</tr>
<tr>
<td>HTR2A</td>
<td>5-Hydroxytryptamine (Serotonin) Receptor 2A</td>
</tr>
<tr>
<td>IBD</td>
<td>Identity by Descent</td>
</tr>
<tr>
<td>JPT</td>
<td>Japanese East Asians</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least Absolute Shrinkage and Selection Operator</td>
</tr>
<tr>
<td>NEO-FFI</td>
<td>NEO- Five Factor Inventory</td>
</tr>
<tr>
<td>NPV</td>
<td>Negative Predictive Value</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
</tr>
<tr>
<td>MDS</td>
<td>Multidimensional Scaling</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forest</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristics</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error</td>
</tr>
<tr>
<td>SES</td>
<td>Socioeconomic Status</td>
</tr>
<tr>
<td>SLE</td>
<td>Stressful Life Events</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphisms</td>
</tr>
<tr>
<td>YRI</td>
<td>Yorubans from Nigeria</td>
</tr>
</tbody>
</table>
List of Tables

**Table 2-1.** Comparison between the four designated ethnicity measures in identifying the number of White Europeans from our sample of 118 patients. ............................................................... 45

**Table 3-1.** Clinical demographics of the study sample based on stratification of suicide attempters and non-attempters. ........................................................................................................ 60

**Table 4-1.** Summary statistics for clinical, demographic, and sociocultural predictor variables. Predictor variable type is indicated as either categorical or continuous. ................................................. 78

**Table 4-2.** Comparison of algorithm performance in classifying individual patients as either suicide attempters or non-attempters. Values are provided as means and standard errors. ........ 87
List of Figures

**Figure 2-1.** Multidimensional Scale (MDS) clustering according to geographical ancestry with reference populations from the HapMap Phase II project. ................................................................. 46

**Figure 2-2.** STRUCTURE analysis for ethnicity of study participants and reference population from HapMap Phase II. 1 = Study participants (n = 118); 2 (green) = North/Western Europeans from Utah (CEU), reference population (n = 60); 3 (blue) = Japanese and Han Chinese Asians (JPT + CHB), reference population (n = 90); 4 (red) = Yorubans from Nigeria (YRI), reference population (n = 60) ........................................................................................................................................ 47

**Figure 2-3.** Confusion matrices representing the performance of both administered scales as compared to MDS-defined ancestry and STRUCTURE-defined ethnicity. 1 = European Caucasian; 0 = Other Ethnicity ........................................................................................................ 49

**Figure 2-4.** Receiver Operating Characteristic (ROC) curves and their corresponding AUC values for each assessment-based scale as compared to both MDS and STRUCTURE-defined ethnicity ......................................................................................................................... 50

**Figure 3-1.** Multidimensional Scale Clustering according to Geographical Ancestry using 292 SNP markers from the HapMap Phase II project. Cut-off points for discordant ethnicity and ancestry are visualized at x = 0.05 to distinguish European Caucasians from those of African ancestry, and y = 0.05 to separate European Caucasians from those of East Asian ancestry...... 62

**Figure 3-2.** STRUCTURE analysis of self-reported ethnicity of study participants and reference population from HapMap. 1 = Study participants (n = 276); 2 (green) = Japanese and Han Chinese Asians (JPT + CHB), reference population (n = 90); 3 (red) = North/Western Europeans from Utah (CEU), reference population (n = 60); 4 (blue) = Yorubans from Nigeria (YRI), reference population (n = 60) ........................................................................................................................................ 63

**Figure 4-1.** Flow chart demonstrating the process by which our machine learning algorithms trained and tested our clinical variables in predicting suicide attempters. Stratified K-fold Cross validation was used to maintain class distributions across all folds; T= Test Data ..................... 77
**Figure 4-2.** ROC curve for LASSO regression model across all stratified folds in cross-validation. A final AUC score is given as the average of all AUC values across all 10 folds. .... 81

**Figure 4-3.** Mean confusion matrix describing performance of the LASSO regression model across all stratified folds from the cross-validation. ................................................................. 82

**Figure 4-4.** Mean confusion matrix describing performance of the RF model across all stratified folds from the cross-validation. ................................................................. 83

**Figure 4-5.** ROC curve for Random Forest model across all stratified folds in cross-validation. A final AUC score is given as the average of all AUC values across all 10 folds. ......................... 84

**Figure 4-6.** Visualization of each predictor variable and their average weight in the LASSO regression model for all 10 cross-validation folds. Each variable’s weight indicates the level of importance in predicting the classification phenotype following regularization. Predictor variables demonstrate a positive, negative, or zero weight due to penalization. ......................... 85

**Figure 4-7.** Visualization of each predictor variable and their average weight in the RF model for all 10 cross-validation folds. Each variable’s weight indicates the level of importance in predicting the classification phenotype following regularization. ..................................................... 86
List of Appendices

A) Appendix A: Self-Report Ethnicity Questionnaire ............................................. 124

B) Appendix B: Grandparent Ethnicity Questionnaire .......................................... 125
Chapter 1

1 Literature Review

1.1 Schizophrenia

1.1.1 Overview and Clinical Features

Schizophrenia is a debilitating mental disorder that is prevalent among the world’s population, affecting approximately 1% of individuals across cultures (van Os & Kapur, 2009). Patients with schizophrenia vary in their presentation of symptoms, though these typically include a combination of positive symptoms (delusions and sensory hallucinations); negative symptoms (flat or blunted emotions, poor motivation); and poor cognitive and social functioning (Goff et al., 2005; Phan & Kreys, 2011; Hovington & Lepage, 2012). Schizophrenia is equally common among males and females, and is believed to be the combined result of both environmental and genetic factors (Xu et al., 2013). The average age of onset for the illness is around late adolescence and early adulthood (American Psychiatric Association, 2013), though males typically develop symptoms slightly earlier than females. Many patients with schizophrenia are unable to maintain employment, social relationships and overall functioning as the illness is a major cause of disability. Patients with schizophrenia reportedly die much earlier than the average population; causing more death than many cancers and physical illnesses, due to suicide and poor lifestyles (Saha, Chant, & McGrath, 2007). Schizophrenia can have severe economic, health, and personal consequences on patients and those around them, especially without access to adequate treatment. It presents itself as a major burden to society through medical costs, required government financial assistance, loss of productivity, and continuous inpatient care. While schizophrenia is less prevalent than more common mental health disorders
such as depression, it is the leading cause of disability compared to all mental health and addiction disorders (Whiteford et al., 2013).

The identification of symptoms relating to schizophrenia generally falls into three specific categories: positive, negative, and cognitive symptoms. Positive symptoms are feelings or behaviours that are typically not present in reality, which include sensory hallucinations, delusions, and disorganized speech. Negative symptoms are behavioural deficits that lead to poor functioning and debilitation, which include anhedonia, a lack of motivation, and blunted affect (Fletcher & Frith, 2009). Cognitive deficits are seen in poor performances on cognitive tasks, whereby patients perform lower than the average population. Cognitive symptoms have been reported to be a stronger indicator of overall functioning in patients with schizophrenia as compared to traditional positive and negative symptoms. The extent of these cognitive deficits predicts the success of an individual patient's treatment regimen and outcome (Green, 2006). Overall, schizophrenia is categorized as a mental syndrome characterized by hallucinations, bizarre delusions, negative symptoms, cognitive deficits and few affective symptoms.

Schizophrenia is generally seen as a highly heterogeneous disorder in which there is a large variability in the manifestation of symptoms. As a result, a diagnosis of schizophrenia in one individual may not always be for the same reasons as another patient. According to the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-V), a diagnosis of schizophrenia requires that two of the following symptoms are present for at least six months with at least one month of active symptoms: delusions; hallucinations; disorganized speech; grossly disorganized or catatonic behaviour; and/or negative symptoms (American Psychiatric Association, 2013). As compared to the DSM-IV TR, changes were made to raise the symptom threshold to require that an individual present at least two of the aforementioned symptoms
instead of one. Furthermore, a diagnosis of schizophrenia is no longer accompanied by specific subtypes (i.e. paranoid, catatonic, disorganized, undifferentiated, and residual). Subtypes were defined according to the prevailing symptom at the time of the psychiatric evaluation, however these were found to be ineffective in facilitating treatment because patients' symptoms would often shift and be categorized as one subtype or another over time. Additionally, many patients demonstrated overlapping subtype symptoms, making it difficult to classify a patient under one subtype exclusively. These symptoms must present themselves for at least 6 months and not be a result of organic causes such as traumatic brain injury or substance abuse/dependence. A strong indicator of future psychosis and schizophrenia is the presence of prodromal (early) symptoms (Larson, Walker, & Compton, 2010). These early warning signs appear 2-3 years (on average) before the onset of schizophrenia. Prodromal symptoms typically include slight psychotic symptoms, poor cognitive function (which may be evident at school or work), and dysregulated mood. The current diagnostic method for schizophrenia recognizes the heterogeneity of the condition and allows clinicians to have greater flexibility with diagnostic and treatment decisions.

1.1.2 Epidemiology

The lifetime incidence and prevalence rates of schizophrenia generally fall around 1% although they vary depending on the diagnostic definition of schizophrenia that is used (Dominguez et al., 2009). Intuitively, a narrower definition of schizophrenia results in lower rates than in cases in which a broad definition is used. When allowing for a broad definition of psychosis including delusional disorder, brief psychotic disorder, and psychotic disorder not otherwise specified, a lifetime rate of schizophrenia and related categories was revealed to be approximately 2.3%. The risk for schizophrenia is largely genetic in which estimates from twin studies suggest a heritability of approximately 80%. Although it is well-established that there is a
genetic association with the illness, identifying the specific genetic mechanisms behind the onset of schizophrenia is largely inconclusive. The heterogeneity in the presentation of schizophrenia, primarily because of the uncertainty of the phenotype investigated, is the most likely reason for inconsequential evidence of genetic factors. Nevertheless, a large-scale genome-wide association study (GWAS) has been successful in pinpointing candidate gene variants that are likely more present in schizophrenia patients compared to controls (Ripke et al., 2014). The high heritability of schizophrenia can also be attributed to environmental factors that moderate gene-environment interactions. This suggests that individuals with the genetic predisposition to develop schizophrenia may only display the onset of symptoms when precipitated by an exposure to stressful environmental factors.

1.1.3 Prognosis and Clinical Management

Schizophrenia is a condition in which the course of the illness remains relatively consistent over the lifetime and may require treatment for the duration of the patient's life (Lewis & Lieberman, 2000). The traditional outlook on schizophrenia is that individuals remain debilitated with little hope of an improved outcome. On the contrary, many patients are able to live independently, although with sustained financial and/or living support. The goal of current treatment regimens is to encourage recovery in which the patient is able to move beyond the circumstances of their mental illness and live a meaningful life.

Upon a diagnosis of schizophrenia, antipsychotic medication is typically prescribed as the first line of treatment for psychosis (Harrow et al., 2014). Antipsychotic medications work by blocking dopamine D2 receptors. The first generation of antipsychotics was discovered in the 1950s to treat psychotic symptoms, although not without unwanted motor side-effects. Within the last two decades, second-generation antipsychotics have been developed which largely treat
psychotic symptoms without inducing motor side-effects (van Os & Kapur, 2009).

Unfortunately, these antipsychotics have proven to be somewhat ineffective in treating the negative and cognitive deficits found in schizophrenia (Leucht et al., 2009; Agid & Remington, 2008). Additionally, these antipsychotics are more likely to lead to metabolic side-effects. A clinician's evaluation of the patient's symptoms is used to establish the type of antipsychotic medication and corresponding dosage prescribed. Antipsychotic dosage is generally adjusted and monitored throughout the course of treatment for schizophrenia patients depending on clinical and functional outcomes (Lally & MacCabe, 2015). In fact, for the most severe cases of psychosis, higher doses have been associated with greater improvement in treating positive and negative symptoms (Aronson, 2008).

For optimal treatment outcome for patients with schizophrenia, antipsychotics alone are insufficient. These medications are most effective when allied with psychological and social support (Zygmunt et al., 2002). Treatment protocols that require both medication and community-case engagement are the most likely scenarios in which remission is possible, particularly when these are implemented during the early stages of the illness (Robinson et al., 1999). One of the most imperative but challenging issues with treatment is ensuring that patients continue to take their medication at the appropriate times. Unfortunately, many patients stop their medication due to the stigma associated with using antipsychotics and because of the decrease in motivation associated with the illness (van Zelst, 2009; Thornicroft et al., 2009; Artaloytia et al., 2006). However, when patients are no longer taking medication as part of their treatment, without the advice of their doctor, they inevitable increase their risk for relapse.
1.2 Impact of Immigration on the Development of Schizophrenia

1.2.1 Incidence of Psychosis among Immigrants

There has been consistent evidence which suggests that the rate of psychosis is elevated among immigrant populations and ethnic minorities worldwide (Coid et al., 2008; Cantor-Graae & Selten, 2005; Bourque et al., 2011). This association has been apparent in studies conducted in the United Kingdom, Netherlands, Canada, United States and Denmark (Fearon & Morgan, 2006; Fearon et al., 2006; Bresnahan et al., 2007; Selten et al., 1997). However, the explanation underlying this phenomenon remains unclear and it has yet to be fully understood why incidence rates of schizophrenia may be higher in multiple diverse groups (McKenzie, Fearon, & Hutchinson, 2008). As far back as the early 20th century it was first reported that all migrating individuals have an increased risk of developing schizophrenia (Odegaard et al., 1932). An elevated risk for schizophrenia among immigrants is however not related to higher rates in the immigrants' native countries (Mahy et al., 1999, Hanoeman et al., 2002; Selten et al., 2005). Large-scale meta-analyses (Cantor-Graae & Selten 2005; Bourque et al., 2011) estimated that immigrants were two to three times more likely to develop schizophrenia compared to their native-born counterparts. This was similarly observed in second generation immigrants (those individuals born in the host country to immigrant parents). Interestingly, there are also instances of elevated rates of bipolar disorder among immigrants (van Os et al., 1996; Lloyd et al., 2005). A large-scale meta-analysis by Swinnen and Selten (2007) demonstrated that the relative risk for bipolar affective disorder was significantly increased for immigrants compared to native-born individuals. These studies suggest that immigration also presents itself as a risk for affective disorders, which are often accompanied by psychotic disorders.
As one of the most ethnically diverse countries in the world, approximately 20% of the population was born in another country (Statistics Canada, 2008a). Visible minority groups make up 16.2% of the Canadian population and are defined by Statistics Canada as non-Aboriginal and non-white in skin colour (Statistics Canada, 2008b). With thousands of new immigrants arriving every year, we are witnessing an increase in the immigrant population in Canada (Dealberto, 2013). Before the 1960s, immigrants mainly originated from Europe; however by 2006 many immigrants were arriving from South and East Asia (Hansson et al., 2012). Hospital records from the 1900s indicate that European immigrants to Canada exhibited higher incidence for psychosis (Smith et al., 2006). Similarly, the rate of hospitalization for psychosis in Ontario today is higher among first generation immigrants (Anderson et al., 2015). Adeponle et al. (2012) found that clinicians were more likely to make a diagnosis of psychosis for immigrant and ethnic minorities over native patient populations. This is largely in part due to the clinician’s failure to acquire sufficient diagnostic information while paying attention to cultural nuances in interpreting illness behaviour.

1.2.2 Difficulty in Acculturation

“Acculturation” is a term used to describe changes that groups and individuals experience upon contact with another culture different from their own. The stress associated with acculturation may be accompanied by a host of negative emotions and behaviours leading to psychotic symptoms (Ratkowska & De Leo, 2013). Immigrants to a country where they are starting anew may experience drastic changes in social and financial status, in addition to potential discrimination. Social factors that are a result of the immigrants’ change in geographical location and their reason for leaving can have drastic psychological implications (Anderson et al., 2015; Bhugra, 2004). An individual’s pre-immigration social vulnerabilities, employable skill set, and self-perception have a large influence on their overall psychological
state. After immigration, negative and positive experiences; difficulties in constructing meaningful relationships; and financial or social support, may further reduce the psychological well-being of immigrants (Bhugra, 2004; Clarke et al., 2008).

Immigrants arriving from societies where community and a sense of belonging are highly valued may experience serious difficulties in adaptation when living in individualistic societies in their new country. This poor adjustment may lead to a lack of adequate social support; disappointment from perceived differences between their expectations pre-immigration and their current reality; and ultimately low self-esteem (Bhugra et al., 2011). A sense of loss for their culture and a sense of guilt for having left their country of origin may lead to poor acculturation among immigrants with them feeling excluded from their new country (Ratkowska & De Leo, 2013). Primarily among the youth, difficulties in adjusting to a new country and its culture can be a confusing and traumatic experience leading to the development of mental health disorders (Ampadu, 2011). In fact, there appears to be an association between difficulties in adjusting to a new culture and a higher risk for mental health issues (Gupta & Bhugra, 2009). Despite attempts to mitigate the stressors associated with acculturation, the act of moving a child from an environment in which they are accustomed, to a new country where its people may differ in their culture, religion, and language may lead to confusion and traumatic experiences that initiate the onset of psychotic symptoms (Pumariega et al., 2005). The increased risk for psychosis is primarily thought to occur as a result of socio-economic difficulties and existential despair that individuals may face post-immigration which may even persist through to second and third generation immigrants (Anderson et al., 2015). This suggests that hardships faced post-immigration such as settlement and cultural assimilation may contribute more to an increased risk for psychosis.
1.2.3 Cultural and Ethnic Communities

The diversity in demographic characteristics associated with immigration and ethnicity all contribute to how an individual adapts to and manages life post-migration (Clarke et al., 2008; Termorshuizen et al., 2015). Investigations on ethnicity in the context of migration demonstrate that immigrants are at an increased risk for developing schizophrenia compared to those of the host population (Bresnahan et al., 2007; Fearon & Morgan, 2006; Veling et al., 2006). On the other hand, immigrants and ethnic minorities who reside among members of their own ethnic group are less at risk for developing schizophrenia than ethnic minorities who live in areas with few of their own ethnic group (Veling et al., 2008). Essentially, the feeling of being different from the majority, or "exception to the norm", due to differences in culture and ethnicity is a major psychological contributor to an increased incidence of schizophrenia (Berg et al., 2014).

Researchers have noted an increased incidence of psychosis among individuals living in areas where they are considered a minority in that population (Veling et al., 2008). Conversely, immigrants who reside in an area with members of their own culture are able to integrate more easily as there are fewer intra-group barriers to social interaction (Hansson et al., 2012). It is suggested that immigrants living in neighbourhoods where their ethnicity is not considered a minority may benefit from increased social support due to cultural similarities, thereby preventing the feeling of social isolation and stress. Furthermore, this may protect against experiencing forms of discrimination as these individuals interact more frequently with members of their own cultural group (Veling et al., 2008). This may lead to a decrease in social stress as they are able to identify with those who hold similar values and experiences within their community. The role of social support can be seen in Ontario where a decrease in the rates of psychotic disorders is observed among immigrants from North and South Europe, and from East
Asia due to the high density of these populations already established in Canada (Anderson et al., 2015). This was also found among Arab speaking individuals living in Wayne, Michigan where the lowest rates of suicide occur in this group as they have the largest minority population of Arab-Americans in the United States (El-Sayed et al., 2011). Therefore, it is evident that a lack of social support in the context of cultural and ethnic sensitivity can lead to an increased risk for psychosis, though this effect may be mitigated by an increase in social support by living amongst ethnic and cultural groups similar to one’s own.

1.2.4 Low Socioeconomic Status

Given that poverty and low socioeconomic status (SES) are common among new immigrants adjusting to a new country, there is well-documented evidence supporting the notion that these disadvantages increase the likelihood of developing psychosis among foreign-born children, Canadian-born children of immigrant parents, and children of non-immigrant parents (Beiser et al., 2002; Cantor-Graae, 2007). Previous studies, however, found that poverty among immigrant children was not substantially able to explain the prevalence of mental health issues. However, these studies notably failed to account for sociological variables such as discrimination and difficulties in acculturation (Beiser et al., 2002). Bresnahan et al. (2007) also compared African Americans and European Caucasians to observe whether a high incidence of schizophrenia among African Americans could be explained by SES. The investigation noted that African Americans were approximately three times more likely to have a diagnosis of schizophrenia. However, these results were inconclusive in explaining the effect of ethnicity and SES on the onset of psychosis due to the limitations in their study design.
1.2.5 Selective Immigration

It has been argued that higher rates of schizophrenia amongst immigrants can be explained by the idea that individuals with a higher vulnerability towards psychosis are more likely to emigrate. This idea of ‘selective immigration’ was first advanced by Odegaard (1932) due to observations that the Norwegian immigrants in the study had a history of poor social adaption in their native country. This explanation, however, has not been validated in further studies of this kind (Selten et al., 2002). Notably, it is expected that the cognitive deficits and negative symptoms that characterize the acute phase of psychosis prior to the onset of schizophrenia are more likely to reduce rates of immigration. In one study, Lundberg et al. (2007) investigated rates of psychosis among future emigrants in Kampala, Uganda. There were ultimately no differences in psychotic and manic symptoms experienced between those planning to emigrate and those with no intention of leaving the country. It is also important to note that Canada’s immigration policy requires that hopeful immigrants undergo a prearrival medical examination. Applicants may not be admitted if they are deemed to be a danger to public safety or health, or a financial burden on health and social services (Laroche, 2000). This may lead individuals with prodromal or full psychotic symptoms to be denied migration status in the country. Nonetheless, this practice emphasizes the selection of healthy immigrants to Canada. Unfortunately, there is comparatively less effort directed at maintaining their health post-migration (Beiser, 2005).

1.2.6 Population Genetics

Given the high heritability of schizophrenia, genetic inheritance has been put forward as an explanation for the increased incidence of psychosis among immigrants. This, however, becomes a controversial topic when considering the role of race and ethnicity (Fernando, 1991). Nevertheless, there is a lack of evidence to support the notion that a higher incidence of
schizophrenia is a result of a greater genetic predisposition or vulnerability among some
immigrant and ethnic groups. Given that the incidence rate of schizophrenia is approximately 1%
worldwide and that no specific ethnic group displays higher rates of psychosis, it is unlikely that
population genetics plays a catalyzing role. If genetic differences in populations did underpin the
reported high rates, it would be expected that rates would be high in the originating countries.
Currently, there is limited evidence to support that the prevalence rate of schizophrenia remains
the same in any country for the same immigrant population. Few studies in particular found that
Caribbean populations differ in incidence rates of schizophrenia depending on their host country.
While there are no major population genetic factors that contribute exclusively to the risk for
psychosis, it remains possible that gene-environment interactions may still be at play, wherein
individuals with a genetic predisposition may develop schizophrenia when they have found
themselves in an extremely stressful situation (i.e. the process of immigrating).

1.3 Suicidal Behaviour

1.3.1 Epidemiology

Suicide, which is defined as the act of deliberately causing one’s own death, is a globally
leading cause of death. Each year suicide leads to the deaths of more than 800,000 individuals
around the world (Nock, Borges, & Ono, 2012; WHO, 2014). Suicide carries a substantial
emotional toll on friends, family members and all those surrounding the victim. It is also a
significant public health challenge due to the economic and human costs associated with suicide.
Suicide ranks as one of the leading causes of death regardless of age, though it is more prevalent
among those between the ages of 15-29, where it is the second most common cause of death
among adolescents (Värmik, 2012; Turecki, 2014). There remains a great deal of prejudice and
stigma surrounding suicide, which prevents individuals who may be thinking about committing
suicide from actively seeking help or treatment. The perception of suicide has been largely influenced by religion, culture and society; particularly among followers of the world's major religions where suicide is considered a sinful act (Burshtein et al., 2016). Although there is growing concern for the prevalence of suicide in societies across the world, there has been relatively less awareness about the toll that suicide may cause on public health.

Studies on epidemiological factors for suicide have identified gender as a contributing factor. It is already well established that there are differences in suicidal behaviour between men and women, regardless of whether or not they have been diagnosed with a mental illness. In fact, in the general population, men die up to four times more often by suicide than females, while females attempt suicide much more frequently (Värnik, 2012; Chang, Gitlin, & Patel, 2011). The reported difference in suicide rates for males and females is partially a result of the methods used by each gender. Although females attempt suicide at a higher rate, they are more likely to use methods that are less immediately lethal (Canetto & Sakinofsky, 1998). Males frequently complete suicide via high mortality actions such as by hanging or firearms. This is in contrast to females, who tend to rely on drug overdosing. This is commonly known as the gender paradox of suicidal behaviour (Canetto & Sakinofsky, 1998). A major exception to this trend is found in China where females have one of the highest suicide rates globally and is the only country where females complete suicide at a higher rate than men (Värnik, 2012). Among the world's populations, suicide rates are similar between men and women in the Eastern Mediterranean, while the highest rate of suicide among females is in South Korea (Värnik, 2012). Notably, the difference in gender suicide rates appears to be less apparent in developing countries where the use of lethal agricultural pesticides by young women living in rural areas is prevalent. It has also been suggested that it may be due to the increased value of motherhood in these cultures in
which the female's identity is based upon the family and the suicidal females having individuals who depend on them, leading to a lower likelihood of suicide (Payne et al., 2008).

1.3.2 Immigration and Suicide

Early studies have found that immigration increases the risk for suicide (Sundaram, Quin, & Zollner, 2006; Westman et al., 2006). Notably, the aforementioned risk factors for schizophrenia associated with immigration are also associated with suicidal behaviour. These include previously described risks associated with acculturation; cultural and ethnic communities; low SES; and population genetics. The stressors associated with emigrating to a new country lead to the development of many psychological issues, including depression, anxiety, psychosis, and suicide. In particular, immigration may lead to significant changes in suicide ideation, frequency of suicide attempts and rates of suicide of people that migrate compared to the host country (Ratkowska & De Leo, 2013).

The act of leaving one’s country of birth is a major life change that is likely to be a stressful situation leading to psychological distress. Many individuals face drastic changes in social roles and social status, and are potentially exposed to social marginalization, prejudice, and discrimination by the host population (Bhugra, 2005; Shoval et al., 2007). Similar to the hypothesis for increased rates of schizophrenia among immigrants, it has been suggested that difficulties in acculturation may explain increased suicide rates (Hovey, 2000). Furthermore, poor employment prospects and a failure to integrate into society are also thought to be driving factors behind increased rates of suicide (Fossion et al., 2004; Garcia & Saewyc, 2007). Particularly, for some ethnic groups there may be a negative perception leading to racial discrimination and prejudice, leading to isolation, loneliness and feeling like an outsider, ultimately resulting in an increase in suicidal behaviour (Mullen & Smyth, 2004).
The process of migration can be understood as consisting of three phases: pre-immigration; the act of emigration; and post-immigration. The experiences and difficulties associated with each phase may vary between individuals and also lead to unique stressors associated with immigration (Bhugra et al., 2011). As we’ve seen with higher incidences of schizophrenia, immigrants may have higher rates of suicidal behaviour compared to individuals in their new country because of the stressors associated with the migrating. The loss of connection between their country of origin and the social network they once had may contribute to these negative emotions. The need to start over after losing their social and financial status comes with a feeling of inadequacy, further propagated by language barriers, unemployment, and a sense of not belonging. This overall feeling of exclusion can lead to an inability to form meaningful relationships and the onset of suicidal behaviours (Iliceto et al., 2013).

Research on the risk for suicide has attempted to look at international rates of suicide among immigrants. Comparisons of suicide rates between immigrants and their host population have resulted in conflicting results when looking across cultures and societies. There have been many trends and findings reported for immigrant suicide research, namely: 1) immigrants originating from countries with low suicide rates tend to display similar rates to the host country (Burvill, 1998; Hjern & Allebeck, 2002; Westman, Sundquist, Johansson et al., 2006); 2) immigrants show higher rates of suicide in the host country compared to what is seen in their country of origin (Merrill & Owens, 1988; Sainsbury & Barraclough, 1968; Whitlock, 1971); 3) immigrant suicide rates are lower than the host population (Crawford et al., 2005; Malenfant, 2004); 4) immigrant suicide rates decrease (Kliewer & Ward, 1988; Sorenson & Golding, 1988) or converge with that of the host population over time in the new country (Greenfield et al., 2006; Singh & Hiatt, 2006). Plausible reasons for this discrepancy have been attributed to cultural differences (Maˇkinen, 2009) or genetic influences (Voracek & Loibl,
2008). Nonetheless, future studies of this kind are required to correctly ascertain the influence of immigration on suicide while taking into account both cultural and genetic factors.

1.3.3 Suicide Terminology and Definitions

As a testament to the complexity, heterogeneity and widespread prevalence of suicide, it was recently proposed that suicidal behaviour should be considered for inclusion in the DSM-V (American Psychiatric Publishing, 2013). Suicidal behaviour is often described on a continuum, ranging from suicidal ideation to non-suicidal self-injuries, suicide attempts, and suicide completion. Suicidal ideation generally refers to when an individual has enduring thoughts or wishes concerning suicide (Gliatto & Rai, 1999). Among this category alone, there is great variability in suicidal ideation which can range from passive thoughts to more extensive detailed planning, with full intent to act on these thoughts. Suicide attempts are further along the continuum in which suicidal thoughts progress into concrete actions. A suicide attempt is defined as an act where individuals make preparations and carry out a suicidal action with greater than zero intent to kill themselves, although the action is unsuccessful (Levi-Belz & Beautrais, 2016). This may also be considered a failed or nonfatal suicide attempt. It is important to distinguish this from an interrupted or aborted attempt in which an individual does not carry out a suicidal action because of external intervention or because of self-restraint, respectively. Lastly, suicide completion is when an individual dies from a suicidal action. There is a great deal of variability between the different categories of suicidal behaviour, which may represent differences in underlying etiology and psychopathology. Suicide research therefore requires the careful examination of suicidal behaviours as stand-alone phenotypes in an attempt to uncover the underlying mechanisms.
1.3.4 Investigating Suicide Attempts

For many years, one of the major methods in which suicide research was conducted was through a retrospective analysis of individuals who died by suicide. This was mainly done by reviewing medical and legal records as well as by speaking with key individuals who may be able to offer relevant information on the deceased individual (psychological autopsies) (Clark & Horton-Deutsch, 1992; Hawton et al., 1998). This approach has been vital in many studies on suicide around the world and is particularly useful for investigating post-mortem brain samples of those who died by suicide. Although this method can reveal vital information regarding suicide, a major limitation to this approach is the lack of access to direct personal information regarding the problems and processes that led to suicide completion (Hawton et al., 1998). This is especially problematic as many family members and close friends may be surprised to know that someone they knew died by suicide (as they did not notice the warning signs), and therefore may not have much insight into their motives or would like to conceal the truth as much as possible because of stigmatization or insurance issues.

An alternative strategy for investigating the mechanisms of suicide is by focusing on those who survive a suicide attempt. A suicide attempt is an act that was likely to result in death but for some reason or another, the individual survived (Beautrais, Joyce, & Mulder, 1999). Individuals who have a history of a serious suicide attempt are important proxies for studying completed suicide as their psychological profile is very similar (Beautrais, 2001). Therefore, this approach has the potential to uncover key mechanisms motivating suicide completers, i.e., whose suicide attempt was ultimately fatal. The main advantage of studying suicide attempt is that researchers have access to individuals who are still alive and thus they can continue to provide first-hand clinical and psychological information on the nature of their suicide attempt. While recall bias is an important problem in this methodology, performing clinical assessments on these
individuals allows us to observe those who were as close as possible to death by suicide and obtain essential data on risk factors and early warning signs (Hawton, 2002). By assessing individuals who have survived a suicide attempt we are able to conduct a detailed investigation into the psychological processes that led to the attempt and the influence of early life experiences, including potential conflicts, decisions prior to the attempt, levels of depression and hopelessness, impulsivity, aggressive feelings, the degree of planning, and even biological characteristics (Beautrais, 2004; Hawton, 2002). Accurately defining the suicide phenotype is essential. In subsequent chapters of this thesis, lifetime suicide attempt history is the main grouping variable in classifying individuals as either suicide attempters or non-attempters.

1.4 Suicide in Schizophrenia

1.4.1 Prevalence and Incidence Rates

Having schizophrenia has been shown to substantially increase an individual’s risk for suicide (Brown, Inskip & Barraclough, 2000). Individuals with schizophrenia generally experience a range of symptoms, such as delusions, sensory hallucinations, flat or blunted emotions, and disorganized thoughts and speech (Phan & Kreys, 2011). These symptoms may be associated with an increased risk for self-injurious and suicidal behaviours (Brown et al., 2000). In particular, suicide attempts and suicide completion are amongst the largest contributors to the increased morbidity and mortality rates in schizophrenia (Hawton et al., 2005). A meta-analysis by Palmer et al. (2005) estimated that one in every twenty individuals with schizophrenia will commit suicide, placing the risk for suicide at approximately 4.9% (Palmer, Pankratz, & Bostwick, 2005). On the other hand, many studies have suggested approximately 20-40% of patients with schizophrenia will attempt suicide during their lifetime (Pompili et al., 2007; Siris, 2001; Altamura et al., 2003; Suokas et al., 2010). Individuals with schizophrenia also tend to
attempt suicide at a younger age than the general population (Limosin et al., 2010). The risk for suicide attempt between male and female patients has been debated, with some studies suggesting that males are more likely to attempt suicide (Harkavy-Friedman, Nelson, & Venarde, 2001) and others claiming that women are at a greater risk than men (Canuso & Pandina, 2007). Most recently, in a clinical sample of patients with schizophrenia and noninstitutionalized controls, it was found that 39.2% of the patients had attempted suicide compared to 2.8% of the controls (Fuller-Thomson & Hollister, 2016). Overall, they found that the schizophrenia patients were six times more likely to attempt suicide compared to those without the illness while controlling for pertinent clinical variables such as depression, anxiety, substance use, childhood adversity, and sociodemographic factors. It has also been suggested that there are similar rates of suicide attempt between individuals with schizophrenia in the community and those who are institutionalized. The current and long-standing goal of suicide research in schizophrenia has been to identify risk factors that may enable clinicians to accurately identify individuals at the greatest risk for suicide. The overarching aims of this thesis are motivated by this important goal.

1.4.2 Risk Factors for Suicide in Schizophrenia

Discrepancies in the ability to identify suicide risk factors in schizophrenia are largely attributed to differences in research methodologies and target samples that attempt to assess single risk factors for suicide. This inconsistency may be attributed to epidemiological studies focusing on patients at different stages of the illness. Notably, first-admission individuals with schizophrenia are at a higher risk for suicidal behaviour compared to chronic patients (Alaraisanen et al., 2009). The risk for suicidal behaviour in schizophrenia is greater during the early years after disease onset, typically during the first 5 years following a diagnosis (Heilä et al., 2005). Despite a relative decline in the suicide rate among older adults with schizophrenia, it is still greater than among the general population (Cohen, Abdallah & Diwan, 2010; Montross et
al., 2008). An improved ability to predict suicide risk in schizophrenia is essential for clinicians to make crucial decisions for therapeutic interventions.

1.4.2.1 Clinical Demographics

Many studies have attempted to identify key demographic traits that are predominantly shared among those with schizophrenia who also demonstrate suicidal behaviour. Among patients with schizophrenia, the age, gender, marital status and employment status of these individuals have all been found to contribute to the likelihood that they will attempt suicide or complete suicide.

It appears that there are two major peaks in age for when suicide risk is highest. Previous studies have demonstrated that being of a younger age group leads to an increased predisposition for suicide, compared to the general population (Hor & Taylor, 2010). Interestingly, completed suicide occurs at a younger age in schizophrenia (at around 30 years old) compared to other psychiatric disorders (Gómez-Durán, Martin-Fumadó & Hurtado-Ruíz, 2012). Another study suggested that there is a higher risk for suicide also among patients over the age of 50 (Hor & Taylor, 2010). Indeed, it appears that there are two major peaks for when suicide risk is highest in schizophrenia; first during young adulthood and again during later life.

Individuals with schizophrenia typically develop symptoms at a younger age. According to van Os and Kapur (2009) the peak period for the onset of schizophrenia is during late adolescence and early adulthood. Additionally, as the risk for suicide also appears to peak around five years following illness onset (Carlborg et al., 2010) it is not surprising that suicide risk is associated with younger age. Particularly, individuals with first-episode psychosis (FEP) with longer duration of first hospital treatment, the presence of depressive symptoms, and nonadherence to treatment in early phases of follow up after FEP predict future suicide attempts.
When diagnosed at a young age, an FEP patient who believes that having schizophrenia may lead to a deteriorating mental state and lower quality of life can react differently to someone with similar high insight but a more optimistic outlook on the course of the disorder (Melle & Barrett, 2012). Several studies indicate that beliefs about psychosis can influence insight and aspects related to suicide risk. Patients with high insight and low levels of stigmatizing beliefs report lower levels of depression, more hope, better self-esteem and better quality of life than patients with high insight and high levels of negative beliefs (Staring et al., 2009; Linden & Godemann, 2007).

Young adulthood is a critical period in an individual's social and career growth, and development (Addington, 2007). The debilitation from the onset of schizophrenia during these formative years may lead to poor employment rates and difficulties in maintaining relationships. These situations are also independent risk factors for suicide in schizophrenia. Previous reports have found that being single, unemployed, and unable to work generally increases the risk for suicide in the overall population and among individuals with schizophrenia (Popovic et al., 2014).

Further studies on demographic risk factors for suicide in schizophrenia have also identified gender as a contributing factor. It is already well established that there are differences in suicidal behaviour between men and women, regardless of whether or not they have been diagnosed with schizophrenia (Carlborg et al., 2010). In fact, in the general population, men die more often by suicide than females, while females attempt suicide much more frequently. In contrast, among the schizophrenia population, women have been described as displaying male tendencies for suicide whereby they act impulsively and aggressively and choose lethal methods for suicide (Caldwell, 1990). The odds ratio for suicide in schizophrenia in men compared to
women is approximately 1.57 (Hawton et al., 2005), suggesting a less extreme gender difference in suicide risk than seen in the general population. Nevertheless, while less frequently reported, other studies have suggested that women have a higher risk for suicide in schizophrenia or that there is no difference in suicidal behaviour between genders (Popovic et al., 2014).

Considering all reported demographic factors that contribute to suicide risk, some studies have attempted to assign a diagnostic profile for the average suicide attempter with schizophrenia. Fenton (2000) characterized a typical patient with schizophrenia at high risk for suicide as a young male, with good pre-morbid functioning and a late age of first hospitalization, a high IQ, categorized under the paranoid subtype, and retention of abstract thinking. While these variables do contribute to the suicide phenotype in schizophrenia, the relative contributions of each in determining suicide risk remains ambiguous. An updated review by Pompili et al. (2007) characterized a high-risk suicide phenotype as typically a young Caucasian male, never married, with good pre-morbid functioning, post-psychotic depression, a history of substance abuse, and previous suicidal behaviour (attempt, self-harm, ideation). Many of these identified risk factors are considered in today's risk assessments; however further investigations have taken into account the role of more recently reported risk factors.

1.4.2.2 Illness-Related Factors

Schizophrenia is a debilitating mental disorder that entails a range of positive and negative symptoms that are often exacerbated by comorbid affective disorders/symptoms, and treatment complications. Given the degree of dysfunction and lifetime disruption that an average individual with schizophrenia endures, it is not surprising that a significant proportion of these individuals attempt or consider suicide.
One of the most debatable illness-related risk factors for suicide in schizophrenia is the impact of a patient’s insight into their illness. Insight is generally described in psychiatry as being able to recognize one's own mental illness and its consequences. Insight can have major repercussions on many related clinical aspects over the course of the illness, such as recognizing the need for treatment and complying with medication recommendations (Amador, 2004). Intact insight in schizophrenia is associated with medication adherence, lower symptomatology, and better treatment outcomes. At the same time, greater insight is linked with hopelessness, depression, and suicide (Amador, 2004). Generally, insight has been proposed to be a factor which increases the risk for suicide in schizophrenia; however, it has not been sufficiently investigated and conflicting results have been reported. In a meta-analysis by Hawton et al. (2005), there was no statistically significant association between the presence of insight and the risk for suicide. This finding, however, is likely attributed to the small sample size of the studies used to investigate insight and suicide risk.

The major argument in favour of insight leading to higher suicide risk in schizophrenia is that it is often correlated with a sense of hopelessness, demoralization, and depression. Furthermore, feeling hopeless is one of the most consistently reported risk factors for suicide in schizophrenia and other psychiatric disorders (Amador, 2004). On the other hand, other studies have found that individual insight is the first step in gaining mastery over their illness and is usually associated with improved treatment compliance (Lopez-Moringo et al., 2012). Notably, Hawton et al. (2005) considered insight and compliance as separate entities and found a significant association for poor compliance as a risk factor for suicide. Therefore, it can also be argued that better insight may also reduce suicide risk via improved treatment compliance.
Whether or not it is accepted that hopelessness, self-depreciation, and depression are direct consequences of clinical insight, it remains clear that these affective symptoms pose a high risk for suicide in individuals with schizophrenia. Hopelessness has often been proposed as an important risk factor for suicide, perhaps even more so than depression (Friedman et al., 2004). Patients who feel hopeless about their situation as a result of their illness and life circumstances are often the ones that resort to suicide. In fact, hopelessness is a key risk factor that intersects with illness insight. Some studies have suggested that insight into the illness is only associated with suicide if it is followed by feelings of hopelessness (Amador, 2004). For that reason, hopelessness can be seen as the deciding factor in the debate as to whether or not insight leads to increased suicide risk. When individuals’ insight into their illness is high but they do not have severe feelings of hopelessness, it is more likely that they will experience the benefits of good insight. On the other hand, if individuals with schizophrenia have clear insight into their illness and see their situation as hopeless, they are then more likely to develop suicidal behaviour.

Depression another major factor in predicting suicide risk. Previous studies have demonstrated that many schizophrenia patients experience a mood disorder during their lifetime (Hawton et al., 2005). In fact, many of these individuals experience depressive symptoms concurrently with suicidal behaviour (Hawton et al., 2005; Pompili et al., 2013). Specifically, having a comorbid depressive disorder is believed to elicit suicidal behaviour in patients with schizophrenia. Unfortunately, many studies have failed to take into account the timing and severity of depressive symptoms in assessing suicide risk (Carlborg et al., 2010). It appears that one of the highest periods of suicide risk occur during the 6 months post-discharge phase following hospitalization, when the illness remains active and during which the relationship between depression and suicide is proposed to be highest (Heilä et al., 2005).
Hospital admissions are an important risk factor for suicide in schizophrenia as they demonstrate important evidence regarding the overall functioning of the patient. Previous findings have suggested that suicide risk is highest when considering the time at which hospitalizations occur (Qin & Nordentoft, 2005). Approximately one-third of schizophrenia and general inpatient suicides occur during admission or one week following hospital discharge. There is also evidence that suggests that the risk for suicide remains high for as long as one year following discharge (Kao & Liu, 2011). As one would expect, the number of hospitalizations is also indicative of the severity of the illness. This may represent the cumulative extent to which an individual experiences any debilitating positive symptoms, leading to previously described feelings of hopelessness and depression.

When assessing the risk for suicide in schizophrenia, it is necessary to examine the extent to which psychotic symptoms associated with the illness play a role. Historically psychosis has been associated with increased suicidal behaviour in schizophrenia (Fenton, 2000; De Hert, McKenzie & Peuskens, 2001). At first glance, this may be explained by the tendency for irrational thinking during periods of psychosis. Despite any initial logical reasons for why someone with schizophrenia would want to attempt suicide, experiences from the content of delusions or thought disorders may be the deciding factor. Another way in which psychotic symptoms may act as risk factors for suicide is that they may drive the patients to inflict self-injury without suicidal intent but their impaired judgment leads them to miscalculate the probability of death from their action (Fenton, 2000). Additionally, these individuals may be experiencing delusions in which they feel that they are impervious to physical harm or they may experience hallucinations that distract them from this risk (Wong et al., 2013). Even though these are plausible explanations for the association between psychosis and suicide risk, there is little substantive evidence for these in the literature.
Another explanation for an increase in suicide risk during psychosis is the idea that individuals respond to command hallucinations that instruct them to kill or harm themselves. Command hallucinations are cases in which patients hear voices which tell them to perform certain acts, occurring in 18-50% of the schizophrenia population, including the command to kill themselves (Wong et al., 2013). However, there is limited and conflicting evidence for command hallucinations being associated with suicide in schizophrenia. In one of the first controlled studies investigating an association between suicidal behaviour and command hallucinations, Hellerstein et al. (1987) found no significant differences for suicide ideation in schizophrenia. On the contrary, however, additional studies have shown an association between having command hallucinations and a history of suicide (Wong et al., 2013).

Previously, studies have suggested that psychotic symptoms may not be the actual cause for an increase in suicide risk in schizophrenia and that it is the distressed and low mood that follows (Hu et al., 1991). This concept has been observed in previous reports in which the more severe levels of psychosis are associated with suicide. This pathway to suicidal behaviour in schizophrenia represents another way in which multiple risk factors interact.

More recent research regarding psychosis and suicide risk in schizophrenia has attempted to observe whether specific symptoms, namely positive and negative symptoms, are more likely to lead to increased suicide risk (Pješčić et al., 2014). Positive symptoms typically include hallucinations, delusions and disorganization, while negative symptoms usually involve flat affect, anhedonia and a lack of motivation. In contrast to analyzing command hallucinations only, studies that have considered all positive symptoms in its risk for suicide have found more promising results (De Hert, McKenzie & Peuskens, 2001). Even though there is little evidence for command hallucinations playing a significant role in suicide risk, several studies have
suggested that active positive symptoms are associated with attempted suicide (De Hert, McKenzie & Peuskens, 2001).

1.4.2.3 Comorbid Substance Abuse

Substance abuse is commonly associated with schizophrenia and is often believed to influence suicide risk in these individuals. In a previous study, comorbid substance abuse was found more often among younger individuals with schizophrenia who exhibited suicidal behaviour (Pompili et al., 2012). While substance abuse is often considered when assessing suicide risk in schizophrenia, there is not necessarily a difference in the frequency of drug abuse between individuals with schizophrenia who attempt and do not attempt suicide (Harkavy-Friedman et al., 2004). Conversely, individuals with schizophrenia who do have substance abuse may experience more positive symptoms, particularly hallucinations, and correspondingly more suicide attempts compared to patients with no substance use (Sokya, 2000). Notably, hallucinations have been found to increase the incidence of suicide attempts in schizophrenia independently from alcohol or drug abuse/dependence.

As one would expect, substance abuse is associated with impulsiveness and thus suicidal behaviour. However, in general this association is not unique to schizophrenia. It appears that substance abuse, while not directly influencing suicide risk, does in fact impact many other risk factors for suicide that may increase the risk for suicidal behaviour in schizophrenia, such as treatment noncompliance, loss of self-control, violence, and economic difficulties (Harkavy-Friedman et al., 2004). These risk factors may thus eventually become stressors which trigger suicidal behaviour later down the road. Given these examples, current research on suicide in schizophrenia takes into account the role of substance abuse, not as an independent factor but as a modulating risk factor that may increase the risk for suicide. While reports of alcohol abuse
and suicide in schizophrenia have been conflicting, a recent study by Zai et al. (2014) demonstrated the interaction between specific genetic variants and suicidal behaviour that were only significant when alcohol abuse or dependence were considered in the analysis.

1.4.2.4 Previous Suicide Attempt

Previous suicide attempts or ideation have a strong positive correlation with completed suicide in schizophrenia patients and have been among the strongest predictors for suicidal behaviour in patients with first-episode psychosis (Hor and Taylor, 2010). Moreover, the strongest predictor for suicide attempt in schizophrenia is a past history of deliberate self-harm behaviour which includes past suicide attempts (Hawton et al., 2005). A history of suicide attempt appears to elevate the risk for completed suicide threefold (Reutfor et al., 2009; McGirr et al., 2008). McGirr et al. (2008) found that individuals with schizophrenia who were admitted for attempted suicide had the highest risk (of all variables studied) for committing suicide. Reutfor et al. (2009) reported that the proportion of suicide patients with an earlier suicide attempt was 32%. This is lower than the 40-71% estimated in previous studies (Heilä et al., 2005). However, Reutfor’s finding of a five-fold higher suicide risk associated with a history of prior suicide attempt is in the same range as several previous findings, and comparable to the risk estimate obtained in the meta-analysis by Hawton et al. (2005).

Patients with a history of suicidal behaviour (attempt, ideation, self-harm) may also be suffering from emotional dysregulation that may be related to affective symptoms or a general inability to express emotion in an empathetic way (Hawton et al., 2005). Suicidal ideation often fluctuates and many patients may think about suicide (McGirr et al., 2008). Although active suicidal ideation is considered a significant risk factor for suicidal behaviour, it may act as a deterrent in some cases (when not associated with a command hallucination) (Nordentoft et al.,
2004). More specifically, it may be a coping factor for schizophrenia patients to engage in thoughts about the emotional significance of suicide and reasons for performing the behaviour. Nevertheless, it is important that clinicians monitor active suicidal ideation in patients that have attempted suicide in the past.

1.4.2.5 Lifetime Stress

Currently, one of the most interesting risk factors for suicide is a history of stressful life events (SLEs), particularly childhood abuse and trauma (Enns et al., 2006). Unfortunately, only few studies have examined SLEs and suicidal behaviour in schizophrenia, and the vast majority of them are association analyses for suicide completion. Stressful life events frequently precede suicide and they were identified in almost half the cases of suicide in schizophrenia in a Finnish national psychological autopsy study (Heilä et al., 1997). Furthermore, another study (Enns et al., 2006) found that suicide in schizophrenia is more likely to occur following a severe life event associated with aggressive behaviour or psychiatric impairment, and a severe life event of a humiliating nature.

In a retrospective cohort study by Melle et al. (2013), life dissatisfaction was found to be a significant predictor of suicide in FEP patients. It is possible that the FEP patients are in a period of their life that is accompanied by high stress, and the burden of a chronic illness may interact with predisposing suicide traits (Fenton, 2000). In a retrospective report by Cohen, Abdallah and Diwan (2010), suicide risk in older adults with schizophrenia was associated with the lifetime traumatic and victimization scale, revealing that lifetime stressful events do contribute to the suicide phenotype. In particular, higher rates of early sexual and physical abuse were modestly associated with suicide attempt history (Cohen, Abdallah & Diwan, 2010).
Roy (2005) analyzed the effects of childhood trauma on 50 chronic schizophrenic patients with and without a history of suicide attempts. The author found that schizophrenic patients who attempted suicide had significantly higher scores for the Childhood Trauma Questionnaire (CTQ) subscales of emotional abuse, physical abuse, sexual abuse, emotional neglect, and physical neglect. The suicide phenotype was also associated with a higher CTQ total score. Despite the clear association between childhood adversity and suicidal behaviour, this association may have been a result of increased psychopathology in schizophrenia patients that attempted suicide (Funahashi et al., 2000). Ucok and Bikmaz (2007) found that childhood trauma was associated with increased psychopathology and suicide attempts before hospital admission in patients with FEP. This finding is interesting because previous studies have found an association between early life stress and development of mental illnesses. Therefore, it is not conclusive that early life stress contributes to the suicide phenotype directly, but rather indirectly through its effects on illness severity (Cohen, Abdallah & Diwan, 2010; Montross et al., 2010).

1.4.2.6 Neurobiology and Genetics

The complexity of schizophrenia is considered to be the result of multiple interacting biopsychosocial factors. In addition to the previously described psychosocial risk factors, there is also evidence for a biological basis for suicide. Familial studies, adoption studies, and twin studies suggest that suicide is both genetic and heritable (Glowinski et al., 2001; Fu et al., 2002). Previous studies also suggest that the heritability of suicidal behaviour is 43% (Bondy, Buettner, & Zill, 2006; McGuffin, Marusic, & Farmer, 2001). Notably, monozygotic twins show a greater concordance rate for attempted and completed suicide compared to dizygotic siblings (Roy & Segal, 2001). Schizophrenia patients who are relatives of individuals who committed suicide have increased risk of attempting suicide likely caused by familial aggregation (Trémeau et al., 2001; Qin et al., 2002). The familial aggregation of suicide and schizophrenia diagnosis also
extends to siblings of schizophrenia patients who have a significantly increased risk for suicide (Björkenstam et al., 2014). For instance, suicide completion was eight times higher in the offspring of fathers with schizophrenia. Furthermore, Ljung et al. (2013) found that the risk for suicide attempt or suicide completion in offspring of schizophrenia parents was two times higher. Given the aforementioned evidence, it is possible that the genetic risk profile for suicidal behaviour in schizophrenia is similar to the genetic risk for developing schizophrenia.

Early attempts at uncovering biomarkers for suicide aimed to evaluate serotonin dysfunction in suicide victim's brains by measuring cerebrospinal fluid (CSF) monoamine metabolites and serotonin receptor-binding in fronto-orbital brain tissue. While low levels of serotonin activity were associated with suicide in the general population, this was not replicated in a schizophrenia population (Chistiakov, Kekelidze & Chekhonin, 2012). Only one study found a significantly lower concentration of the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA) in the CSF of individuals with schizophrenia who completed suicide (Mann, 1999). While the serotonergic system is one of the most extensively studied in terms of suicidal behaviour, inconclusive evidence for several serotonin receptors and candidate genes has been widely reported (Anguelova et al., 2003; Judy et al., 2012; Buttenschøn et al., 2013).

Continuing onward, many studies began looking for potential genes or genetic markers that might influence suicidal behaviour in schizophrenia. A large portion of these studies have been candidate gene studies in an attempt to identify specific genetic variants or single nucleotide polymorphisms (SNPs) that may indicate a risk for suicide. The most investigated genes believed to be associated with suicidal behaviour in schizophrenia are those of the serotonergic system. One of these major candidate genes is the serotonin 2A receptor (HTR2A) (Serretti, Drago & De Ronchi, 2007). More than 200 SNPs along the gene have been characterized, but only two SNPs (rs6313 (T102C), rs6311 (A1438G) have been extensively
investigated in suicidal behaviour (Tsai, Hong & Liou, 2011). However, it has been repeatedly shown that rs6313 may not be involved in suicide among schizophrenia patients. Both Ertugrul et al. (2004) and Tan et al. (2002) did not find an association with the T102C polymorphism and suicidal behaviour.

In addition to HTR2A, tryptophan hydroxylase (TPH) genes were among the first serotonergic genes to be investigated in regards to suicidal behaviour, specifically TPH1 and TPH2. An early study by Paik et al. (2000) examined a Korean sample of individuals with schizophrenia who had and had not attempted suicide, in addition to controls. In this study they looked at the A218C variant in the gene TPH1. They found that the C allele was more frequently observed among individuals with schizophrenia who attempted suicide. In a later study by Saetre et al. (2010), A218C as well as A779C were tested among a sample of schizophrenia patients with and without a history of suicide attempt. This study, however, found no significant associations for both polymorphisms when comparing attempters and non-attempter patients. A polymorphism in the promoter region of TPH2 was also examined for a relationship with suicidal behaviour in schizophrenia (De Luca et al., 2005). TPH2 has one polymorphism (−473T>A) in the putative promoter and many highly polymorphic variants in introns within the gene. Allele and haplotype frequencies for this promoter polymorphism and other variants demonstrated no significant difference among a cohort of schizophrenia patient with and without a history of suicidal behaviour (De Luca et al., 2005).

1.5 Computational Psychiatry

A major issue in psychiatry is the difficulty in obtaining a consensus on diagnostic criteria and clearly defining the phenotype or illness of interest (Kapur, Phillips, & Insel, 2012). This is largely attributed to the complexity of the problems faced in psychiatry. Not surprisingly,
Mental health disorders and behaviours are often difficult to classify due to their being a product of the brain; an organ that we still do not fully understand (Huys, Maia, & Frank, 2016). This is particularly evident when classifying suicidal behaviour. Because of the range of behaviours and specific nuances that describe each, it is a difficult phenotype to categorize. Additionally, schizophrenia is an example of a highly heterogeneous condition which is often difficult to diagnose. Mental health disorders are often a result of a combination of genetic and environmental factors, which cannot be easily defined for each individual. In order to advance our understanding in psychiatry and improve upon these classifications it is imperative that quantitative measures be used to synthesize and tease out important information from the large amounts of data that are being produced. Advances made in technology and analytical ability have allowed researchers to now ask questions they never could have expected answers for before, and push the boundaries of our scientific knowledge. Computational psychiatry, as a discipline and a tool, will allow us to transcend what we thought was possible in psychiatry research and develop models for mental health that can have a positive clinical impact.

### 1.5.1 Machine Learning: Classification

Machine learning is a field of computer science concerned with the application of complex algorithms for pattern recognition and computational learning. The application of machine learning to clinical practice is becoming increasingly commonplace. Machine learning approaches have been used to tackle complex problems such as improving diagnostic capabilities, predicting treatment outcomes, and predicting the course of disease progression. Classification in machine learning is a problem in which an algorithm attempts to identify to which class (i.e. label) a set of observations (i.e. sample participants) belong to according to a training sample of data that contains observations with known classes. This is a classic example of pattern recognition or 'supervised learning'. Current techniques for classification in psychiatry
often use data given by symptoms, clinical variables, genetics, or imaging data in order to improve diagnostic selection. On average, machine learning models in psychiatry have often attempted to classify individuals as either patients or controls (Huys, Maia, & Frank, 2016). As an example, one group managed to use neuroimaging data to accurately classify individuals as patients with psychosis or healthy controls (Squarcina et al., 2015). A major concern for these models, however, is the issue of overfitting. This is defined as when your algorithm is able to clearly learn from the provided training data but unable to generalize to new unseen data. Overfitting increasingly becomes an issue with a low number of patients and a high number of predictor variables. One way in which this is avoided is by decreasing the number of predictor variables in the model through regularization. This is done by including a penalization term for many predictors such that weakly associated variables will have a much lower relative statistical weight (Huys, Maia, & Frank, 2016). Cross-validation is another technique which aims to minimize overfitting by separating the sample data into two subsets: training data (used to estimate prediction parameters) and validation (testing) data to the ability of the initial parameters to make classification predictions on different data (Kohavi, 1995). Creating these split datasets allows for the prediction of parameters on new data but also with the disadvantage of decreasing sample sizes. Different cross-validation techniques exist to optimize the algorithm's prediction.

1.6 Overall Summary

Suicide is a devastating outcome for many individuals across the world and is a major contributor to the loss of life around the world (Nock, Borges, & Ono 2012). Schizophrenia is associated with a significantly increased risk for suicidal behaviours (Brown et al., 2000). It has been a consistent issue for clinicians to accurately identify individuals who are at the greatest
risk for attempting suicide. Unfortunately many clinicians fail to recognize individuals at the highest risk for suicide attempt due to a lack of reliable suicide risk assessment tools. It has been suggested that sociocultural factors associated with immigration and ethnicity are important contributors to an increased risk of developing schizophrenia (Li et al., 2012; Cantor-Graae, 2007). The act of leaving one's country of origin to begin a new life in an unfamiliar country is a stressful situation for many individuals. This scenario often leads to psychological distress which may accordingly result in suicidal behaviour. There are many stressors associated with immigration and being of an ethnic minority group, which include severe changes to social status, racial discrimination and prejudice, and poor socioeconomic status (Bhugra, 2005; Shoval et al., 2007). Individually, studies have shown that immigration and related sociocultural factors increase the risk of developing schizophrenia and suicidal behaviour (Berg et al., 2014; Spallek et al., 2015). Therefore, it is possible that the stressors associated with immigration and ethnicity contributes to the risk for suicide, specifically among patients with schizophrenia. Immigrant and ethnic differences in this population may be an important guide for assessing the risk for suicide as they are potential precursors to attempted suicide.

1.7 Study Objectives and Hypotheses

The primary objective of this thesis is to investigate the roles of immigration and ethnicity in predicting the risk for suicide attempts among individuals with schizophrenia. The prevailing hypothesis is that immigration and ethnic minority status are important clinical predictors that increase the likelihood of a patient with schizophrenia in attempting suicide during the course of their illness. This is investigated in multiple parts of this thesis by clearly identifying suicide attempt status in a sample of schizophrenia patients and assessing whether a history of immigration and being of an ethnic minority status (as confirmed by self-report and
genetic analyses) are predictive of past suicidal attempts. We attempted to validate self-report measures to accurately assess ethnicity in our sample. Next, we investigated whether migration and ethnicity are associated with suicide attempt in schizophrenia. Finally, we aimed to incorporate many clinical variables suggested to be associated with immigration and ethnicity, which may precipitate a risk for suicide, into machine learning models to facilitate the improved diagnostic classification of suicide attempters and non-attempters.
Chapter 2

2 Validating an Assessment-Based Measure for Ethnicity Using Genomic Markers

2.1 Abstract

Objective: Population stratification is a confounding factor in genetic association studies due to ethnic differences in allele frequencies. Assessment-based measures are the most common method of collecting ethnicity for large-scale genomic studies. However, these measures may be inaccurate and may lead to genotyping individuals who are later excluded. The current study aims to validate whether an assessment-based measure acts as a reliable tool for genetic ancestry.

Methods: We chose to validate both a self-report measure, which is a standard multiple choice questionnaire in which participants select their ethnicity from a predetermined list, and an interview-based measure that collects the country-of-origin of the participants’ maternal and paternal grandparents. The validity of these measures was tested by comparing their responses to the ethnicity defined by Multidimensional Scaling (MDS) and confirmed using STRUCTURE.

Results: Our grandparent questionnaire was the most consistent with MDS ancestry, predicting European ancestry with 98.3% accuracy and 97.6% sensitivity, and an AUC of 99%. Overall, the performance of the grandparent questionnaire was superior to our standard multiple choice questionnaire in correctly identifying European ancestry in our sample, even when compared to STRUCTURE-defined ethnicity.

Conclusion: Our study suggests that collecting information on the geographical ancestry of each individual’s grandparents provides a more comprehensive view of ethnicity to prevent population stratification and wasted resources on genotyping.
2.2 Introduction

If an association exists for one ethnic group over another for a particular phenotype, then any genetic or environmental risk factor that varies between the groups will also appear to be related to the trait (Wacholder et al., 2000). This phenomenon refers to population stratification, in which cases and controls demonstrate different allele frequencies due to ethnic ancestry rather than a genetic association with a trait (Freedman et al., 2004). These systematic differences in ancestry often lead to false positive findings in genetic association studies. The effects of stratification are argued to be eliminated by matching cases and controls for self-reported ethnicity and geographical ancestry (Wacholder et al., 2002). However, the confounding effect of stratification varies in proportion to the number of samples in the study. Even a small degree of population stratification can impact a genome-wide association study (GWAS) due to the large sample sizes required to detect common variants of a small effect on a complex trait (Anderson et al., 2010). In order to avoid these false positive associations, statistical methods have been applied to identify and correct for these differences.

Despite designing a case-control study that attempts to sample individuals from the same population, hidden fine-scale genetic substructure in the population or inadvertent inclusion of individuals from another population may lead to confounding results (Anderson et al., 2010). Confounding occurs when the population substructure is not equally distributed between cases and controls. Generally, there are three types of population structures that might occur in genetic association studies (Li et al., 2010). Discrete population structures include populations that are easily discernible whereby individuals are clearly separated. Admixed population structure consists of individuals of mixed ancestry to varying degrees making it difficult for these individuals to be separated into discrete clusters. Finally, hierarchical population structure
consists of both discrete and admixed population structures which may be found in multi-ethnic cohort studies (Serre et al., 2008).

Given the increasing problem of population stratification induced by underlying genetic substructures, various methods have been applied to correct for this confounding effect. An early approach included Genomic Control (Devlin et al., 1999; Reich et al., 2001). The Genomic Control approach corrects for population stratification by adjusting association statistics with an overall inflation factor obtained from a set of random markers that are not associated with the phenotypes of interest (Price et al., 2006). The method examines the distribution of association statistics ($X^2$) between unlinked genetic variants typed in cases and controls (Freedman et al., 2004). The association statistic at a candidate allele can be compared with the genome-wide distribution of association statistics for alleles that are probably unrelated to disease; determining whether the candidate allele stands out. If no stratification is evident, the association between unlinked genetic variants and disease should follow a $X^2$ distribution with 1 degree of freedom. In cases of stratification, the distribution of association statistics should be inflated by a value termed $\lambda$, which becomes larger with increasing sample size (Devlin et al., 1999; Reich et al., 2001). Some markers however, differ in their allele frequencies across ancestral populations more than others. Certain SNPs show very different allele frequencies in different ethnic groups, whereas other SNPs do not. Therefore, applying genomic control can result in over-correction for some SNPs and under-corrections for others (Price et al., 2006). As a result, Genomic Control may be insufficient for markers that are very ancestry-informative and redundant for markers with poor ancestry differentiation, which would lead to an overall loss of power.

As a result of increasingly large datasets, population stratification has been assessed using various approaches including structured association tests and dimensional reduction.
methods. The structured method generally categorizes individuals into discrete subpopulations using ancestry-informative markers that are associated with each subpopulation (Pritchard et al., 1999). Therefore, each individual is assigned an association statistic that is calculated to signify the relative composition of each subpopulation group. Dimensional reduction methods include principal component analyses (PCA) or multidimensional scaling (MDS) approaches. Both are multivariate statistical methods used to produce a number of uncorrelated variables (principal components) from a data matrix containing observations across a number of potentially correlated variables (Anderson et al., 2010). The components are calculated so that the first principal component accounts for as much variation in the data as possible in a single component, followed by the second, and so on. For each individual, principal components are calculated to allow the clustering of study samples according to ethnic origin. Both methods are currently the preferred method of choice for examining population stratification in genome-wide association studies.

While methods to identify differences in association based on population substructure continue to be refined and developed for large scale genome-wide studies, the validation of ancestry can unfortunately only be done after significant effort and resources have been invested to collect the data. Much of the recruitment of individuals for these large scale genetic studies relies on self-reported ethnicity from individual study participants. However, there are many cases in which participant testimonials may not be reliable or may lack sufficient knowledge on their ancestors’ ethnic origin. This is increasingly the case for samples collected from a psychiatric population, such as schizophrenia, where patient testimonials may not always be reliable or accurate. Self-reported ethnicity may be biased towards more cultural identity or delusional beliefs about ethnic origin in these samples. Only after participants are recruited, assessed, and genotyped do researchers find out through population stratification measures that
the participant may actually have to be excluded due to them not belonging to a homogenous ethnic population. As a result, investigators who wish to collect a large-scale dataset from individuals in the psychiatric population would benefit from a reliable interview-based approach of collecting ancestry information prior to genotyping. In the current study, we aim to validate the efficacy of a self-report and interview-based measure of ethnicity by assessing whether their responses were consistent with ethnicity defined by the MDS and STRUCTURE approach.

2.3 Materials and Methods

2.3.1 Study Participants

Our analysis included 118 participants recruited from the Centre for Addiction and Mental Health (CAMH) with a diagnosis of schizophrenia or schizoaffective disorder according to the structured clinical interview for DSM-IV (SCID-I/P). Exclusion criteria included observed intellectual disability and the presence of neurodegenerative disorders. Written informed consent was obtained for participation in the study. Clinical information was collected for each participant through a cross-sectional assessment using the structured clinical interview procedure and self-report questionnaires.

We assessed each participant’s self-identified ethnicity using the standard multiple choice questionnaire form (see Appendix A) used at CAMH. This form consists of a variety of options for self-defined ethnicity in which participants are asked to circle the response that best represents their identity. Additionally, we identified ethnicity according to participants’ responses regarding the ancestry of each of their four grandparents using a form we developed in our lab (see Appendix B). This form consists of fewer options than our multiple choice questionnaire though it allows for more specific identification of ancestry. Our primary filter for the final sample was to include participants only with both completed self-reported ethnicity
questionnaires and grandparent ancestry information. We chose to carry out our analysis by categorizing individuals as either Europeans or Other ethnicity, mainly due to the current trend of incorporating only Europeans in large-scale genomics analyses.

2.3.2 Multidimensional Scaling Analysis

The multidimensional scaling (MDS) calculation is a mathematical method to reduce higher dimensional data (each individual genotype) into lower dimensions based on patterns of identity by descent (IBD) relationships (Tian et al., 2008). The MDS requires that a pair-wise IBD matrix be constructed and is therefore more computationally complex. The MDS clustering approach attempts to identify both discrete and admixed patterns of genetic variation and correct for their potential confounding effects by adjusting the position of each subject along identified axes of genetic variation and the cluster membership simultaneously (Li et al., 2008). This model is built using pruned genome-wide genotype data from populations of known ancestry as a reference (i.e. reference genotypes from the HapMap project).

Participants were genotyped using an Illumina-2.5 Omni SNP array and standard quality control filtering was applied. Individuals with missing genotype rates of ≥5% were excluded from the analysis and markers with missing data rates ≥5% were excluded. The MDS approach was done using PLINK v1.07 (Purcell et al., 2007) to correct for ethnic stratification and confirm the selection of individuals in our sample on the basis of white European ancestry. We used 51,194 SNPs which were found in both our genotyped sample and the HapMap Phase II project (Frazer et al., 2007). Three reference populations were used from the HapMap sample: European Caucasians (North/Western Europeans from Utah [CEU]), East Asians (Han Chinese and Japanese individuals [CHB + JPT]), and Africans from Nigeria (Yoruba from Nigeria [YRI]). The second principal component from our MDS was used as the stratifying component to
separate Europeans from non-Europeans. Individuals who were within six standard deviations (Price et al., 2006) from the mean of the CEU HapMap reference sample for the first principal component were classified as European for our analysis.

2.3.3 STRUCTURE Analysis

To confirm whether genetic ancestry is most consistent with either the self-report or grandparent ethnicity form, we also used the program STRUCTURE (Pritchard et al., 2000). Using the Structured Association approach, samples are assigned to subpopulation clusters (K) using a model-based clustering program. This program assigns study samples into discrete subpopulations and then aggregates evidence of association within each subpopulation. For each individual, an association statistic is calculated providing the relative composition of each subpopulation group.

Geographical ancestry was assessed using STRUCTURE (v.2.3.4) to identify population substructure and estimate the relative ethnicity of our sample in order to confirm the results of our MDS analysis. The three HapMap Phase II populations were also included as the reference populations. We classified ethnicity in our sample as European Caucasian or Other. In STRUCTURE, the output provides the relative proportion (%) that each participant is of European, Asian, and African ancestry. As the HapMap European reference population was at least 80% European, individuals in our sample of ≥ 80% European ancestry were considered Europeans. We considered an admixture model, which assumes that individuals may have mixed ethnicity from each of the specified populations (k). The admixture model is a reasonably flexible model for dealing with real populations that may not be ancestrally isolated (Pritchard et al., 2000). We set our number of populations, k, to four groups so as to specify the three HapMap reference populations and our study sample. Our analysis was carried out using 10,656 ancestry-informative SNPs that were randomly selected from all chromosomes from our original
The number of SNPs was selected based on the approximate number that STRUCTURE can handle without the program crashing. Given the large number of SNPs used in the analysis, we ran our model considering a burnin length of 1000. Furthermore, we chose 1000 as the number of Markov chain Monte Carlo (MCMC) reps after burnin in order to attain accurate parameter estimates.

### 2.3.4 Performance Analysis

In order to identify the validity of our administered scales in identifying ethnicity as defined by genetic ancestry, we calculated various performance statistics to test their classification ability. Namely, we assessed the ability of our scales to correctly identify individuals who are European Caucasian or Other Ethnicity. We calculated the sensitivity (true positive rate) and specificity (false positive rate) of our scales and also plotted these as Receiver Operating Characteristic (ROC) curves. Confusion matrices were generated to visualize the number of positive and negative classifications for each scale against our STRUCTURE and MDS analysis (Figure 2-3). The corresponding Area Under the Curve (AUC) values for each ROC curve were calculated as well to characterize the performance of our two scales. All performance statistics were calculated using R v. 3.3.0 and the package "ROCR" to plot ROC curves and calculate their corresponding AUC values.

### 2.4 Results

The number of Europeans versus other ethnicities identified using the four methods described in this paper are summarized in Table 2-1. Using our grandparent ethnicity form, we identified 81 individuals with all four grandparents of European ancestry and 37 individuals of Other Ethnicity. We considered individuals as European only if all four grandparents were of a homogenous European background. Given our self-report measure of ethnicity, we were able to
identify a subset of 78 individuals that self-identified as White European/White North American compared to 40 individuals that selected other ethnic backgrounds.

**Table 2-1.** Comparison between the four designated ethnicity measures in identifying the number of White Europeans from our sample of 118 patients.

<table>
<thead>
<tr>
<th>Population Stratification Method</th>
<th>Number of White Europeans</th>
<th>Number of Other Ethnic Origins</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS MDS</td>
<td>83</td>
<td>35</td>
</tr>
<tr>
<td>STRUCTURE (80% CEU)</td>
<td>81</td>
<td>37</td>
</tr>
<tr>
<td>Grandparent Questionnaire</td>
<td>81</td>
<td>37</td>
</tr>
<tr>
<td>Self-Report Questionnaire</td>
<td>78</td>
<td>40</td>
</tr>
</tbody>
</table>

Using the MDS method, we plotted all individuals according to ancestry relative to the HapMap populations (Figure 2-1). Based on this stratification, we categorized 83 individuals as European and 35 individuals of other ethnicities. These 83 individuals clustered within six standard deviations from the mean of the CEU reference population for the stratifying component. The STRUCTURE plot (Figure 2-2) provides the relative ethnic proportion for each individual as a fraction that each individual was of CEU, CHB+JPT, and/or YRI ancestry. As our analysis classified ethnicity as European Caucasian or Other, individuals from our study sample that had a relative European ancestry proportion of ≥ 80% were considered Europeans to prevent ambiguity in ancestry due to large overlap in ancestry-informative markers among mixed, Middle Eastern, or East Indian individuals. From this classification, we reported 77 individuals classified as Europeans and 41 of other ethnic compositions.
Figure 2-1. Multidimensional Scale (MDS) clustering according to geographical ancestry with reference populations from the HapMap Phase II project.

From our analysis, the MDS and the grandparent questionnaire were the closest in consistency in identifying the number of Europeans. Specifically, the MDS identified 83 European individuals and the grandparent questionnaire identified 81, with two individuals being discordant in their reported ethnicity. Upon closer inspection of the individuals selected by each method, we attempted to identify why certain individuals were captured as European in the MDS but not the self-report or grandparent questionnaire. One individual selected by MDS had one grandparent of Native American ancestry and thus was not counted as European according to the grandparent questionnaire. A second individual selected by the MDS had a grandparent of half African ancestry who again was excluded based on our grandparent questionnaire. Classification
as European according to our grandparent questionnaire is contingent on all grandparents being of European ancestry.

Our self-report measure identified 78 individuals as European (or White North American) compared to the MDS which identified 83 European individuals. In this case, there were six individuals that were discordant in their reported ethnicity. Inspection of the individuals who were identified as European in the MDS but not by the multiple choice self-report questionnaire revealed instances in which participants’ answers may have been confused due to poor cognitive functioning or psychosis associated with schizophrenia. Two individuals reported being of mixed background; however, their background according to the grandparent questionnaire revealed that the mix was from Europe. One individual was unable to answer the self-report and selected ‘unknown’ as their response. Furthermore, another individual selected two answers, White European and Aboriginal. Finally, the same individual with one Native American grandparent

**Figure 2-2.** STRUCTURE analysis for ethnicity of study participants and reference population from HapMap Phase II. 1 = Study participants (n = 118); 2 (green) = North/Western Europeans from Utah (CEU), reference population (n = 60); 3 (blue) = Japanese and Han Chinese Asians (JPT + CHB), reference population (n = 90); 4 (red) = Yorubans from Nigeria (YRI), reference population (n = 60).
selected mixed background, which is accurate but inconsequential in that they did not stratify outside of the European cluster in the MDS plot.

Based on the aforementioned findings, the grandparent ethnicity questionnaire appears to be the most reliable measure for ethnicity according to genetic ancestry. We chose to further validate this finding by comparing the grandparent and self-report form to ancestry defined by STRUCTURE. We found that there was complete concordance between the number of Europeans identified by our grandparent questionnaire and STRUCTURE, in which 81 individuals were classified as European. Comparatively, our self-report questionnaire identified 78 individuals as European, leaving three individuals with discordant ethnicity. In summary, between our two ethnicity forms and the genetically-derived ancestry methods we employed, there appears to be the most concordance when using the grandparent ethnicity form. Specifically, the grandparent questionnaire identifies the same individuals as European most consistently with the MDS and STRUCTURE approach, compared to the self-report ethnicity questionnaire. The grandparent questionnaire displays the least discordance between the two genetic methods (MDS = 2; STRUCTURE = 0).

The performance of our scales in correctly identifying individuals according to our binary classification as either European Caucasians and Other Ethnicity was determined by calculating the accuracy, sensitivity, and specificity of our scales against MDS- and STRUCTURE-determined ethnicity. When analyzing the validity of our grandparent questionnaire, we found an accuracy of 98.3% with a corresponding sensitivity of 97.5% and 100% specificity compared to our MDS analysis. Compared to STRUCTURE ethnicity, we found an accuracy of 96.6%, and a sensitivity and specificity of 97.5% and 94.6%, respectively. When testing the validity of our self-report scale against the MDS, we found an overall lower accuracy at 95.8% with a
sensitivity of 94.0% and specificity of 100%. Lastly, compared to STRUCUTRE, the self-report scale demonstrated an accuracy of 94.1%, sensitivity of 93.8%, and specificity of 94.6%. As a further testament to the performance of our assessment-based scales, our plotted ROC curves (Figure 2-4) and corresponding AUC values provide strong evidence for the performance of our scales in correctly identifying European ancestry in our sample. Based on these performance metrics, it is evident that our grandparent questionnaire demonstrates a superior ability to identify European Caucasians to a similar degree as what would be found using genetic population stratification methods.

**Figure 2-3.** Confusion matrices representing the performance of both administered scales as compared to MDS-defined ancestry and STRUCTURE-defined ethnicity. 1 = European Caucasian; 0 = Other Ethnicity.
Figure 2-4. Receiver Operating Characteristic (ROC) curves and their corresponding AUC values for each assessment-based scale as compared to both MDS and STRUCTURE-defined ethnicity.

2.5 Discussion

In the current study, we aimed to identify which of our ethnicity forms is the most reliable predictor of genetic ancestry as defined by the MDS and confirmed using STRUCTURE. As genomic studies rely on large scale analyses of hundreds to thousands of participants, reliable controls for population stratification must be followed. Not accounting for ancestry differences among sample individuals can lead to false positive or spurious associations. Unfortunately,
using genetic markers to control for ancestry leading to subsequent exclusion of individuals based on ethnicity proves to be a financially burdensome and time consuming practice.

Our analysis demonstrates that collecting the ethnicity of the individuals’ four grandparents is most consistent with MDS-defined ethnicity and proves to be a more reliable method compared to traditional multiple choice questionnaires of ethnicity. The grandparent questionnaire had the highest overall accuracy at 98.3% and sensitivity at 97.6% when compared to the MDS. This indicates a greater degree of overlap in classification of European Caucasians using this scale as compared to the MDS, which is the common method of identifying population stratification in large-scale genomic studies. This method may be more stringent in its selection criteria as all four grandparents must be of a European background to be categorized as European, however, individuals may still lie within the European MDS cluster even with a non-European grandparent. We observed that 83 individuals were considered European in our MDS as opposed to 81 from our grandparent questionnaire. The discrepancy in the 2 excluded individuals comes from the fact that both had a single grandparent of non-European ancestry. This reveals a potential complication in using the ancestry of all four grandparents to determine ancestry. By defining individuals as European only if all four grandparents were European as well, we effectively excluded individuals that still clustered among the Europeans according to the MDS. Nevertheless, while adhering to this strict classification of European ancestry may unnecessarily exclude individuals, it may not always be the case where an individual will be within the defined boundaries of the European cluster. Ultimately, the clustering will then depend on the ancestry-informative alleles transmitted from the non-European grandparent.

While the PCA and MDS are both valid methods of correcting for population stratification in genomic association studies, we chose to use the MDS to correct for genome-wide population
stratification as opposed to the PCA. Both techniques provide similar results, however the MDS has some minor advantages. First, the MDS is a more flexible method compared with the PCA (Tian et al., 2008). A PCA requires that underlying data follow a multivariate normal distribution, while MDS does not require the same restriction. Second, PCA requires computation of a covariance matrix first. The MDS however, can be applied to any kind of distances or similarities. The traditional pairwise IBS distance is only one example of a distance measure to which MDS can be applied.

In traditional GWAS analyses, STRUCTURE is often not a preferred method of choice for parsing out population stratification effects. Primarily, STRUCTURE is computationally intensive and difficult to apply using a large number of SNPs and subjects. In addition, the assignment of subjects to clusters is sensitive to the number of clusters estimated, which is often uncertain and ill-defined (Setakis et al., 2006). Unfortunately, STRUCTURE is inefficient for large-scale genomic analyses. It requires a significant number of permutations to accurately define clusters, which are preemptively decided, and demands a high computational cost. However, STRUCTURE may prove to be useful with a substantially smaller number of SNPs in cases where the autosomal markers display strong ancestry-informative information. Nonetheless, with significantly fewer SNPs, STRUCTURE was able to reliably identify individuals as European most similarly to our grandparent questionnaire.

Given that our sample consisted of individuals with schizophrenia spectrum disorders, it is very likely that scenarios in which participants were confused or delusional occurred during self-report ethnicity measures. Our grandparent questionnaire offers a protective approach against ambiguity or confusion by asking solely for the geographical ancestry of the individual’s grandparents. By naming the country that their grandparents originated from, it is simple for
researchers to then classify individuals as either European or of another ethnic origin. Unfortunately, typical multiple choice self-report measures of ethnicity allow more room for error. First, self-report questionnaires generally provide a limited set of options for individuals to select that may not always be clear to the individual. Second, by allowing the individual to choose the ethnicity they identify with, the possibility that cultural influences may have an impact is more likely. Whether this was a case of delusional ideation or not, this is unlikely to be an issue when using the grandparent questionnaire. Ultimately, while the discordance between the self-report and grandparent questionnaire is only marginally different when compared to the genetically-determined ancestry, even if only the ethnicity of a small number of individuals are correctly identified before genotyping is done, this will still provide researchers with the benefit of saving hundreds of dollars on genotyping costs.

2.6 Conclusion

Our study finds that collecting the geographical ancestry for all four grandparents of each participant is highly consistent with ethnicity defined according to genetic ancestry. The MDS method is one of the most widely used in genome-wide association studies to prevent spurious results due to population stratification. By collecting the ancestry of participants’ grandparents, investigators may avoid the unnecessary waste of time and resources in recruiting participants for large-scale genetics studies and genotyping these individuals; only for them to be excluded in the downstream analysis due to population stratification. As opposed to collecting self-reported ethnicity from psychiatric patients, collecting the geographical ancestry of participants’ grandparents offers a cheaper and more accurate alternative to prevent wasted financial resources on genotyping and excluding those individuals after stratification testing.
Chapter 3

3 Assessing the Risk for Suicide Attempt in Schizophrenia According to Migration, Ethnicity and Geographical Ancestry

3.1 Abstract

**Background:** Suicide is a leading cause of mortality among those afflicted with schizophrenia. Previous studies have demonstrated that the stressors associated with immigration may lead to an onset of schizophrenia and suicide independently in susceptible individuals. However, no studies have shown whether migration may lead to suicidal behaviour for individuals with schizophrenia. Our study proposes that an individual’s geographical ancestry, ethnicity or migration status may be predictive of suicide risk in schizophrenia.

**Methods:** In a sample of 276 participants with schizophrenia spectrum disorders, we conducted clinical assessments to collect self-identified ethnicity, migration history, and suicide attempt history. Self-identified ethnicity and suicide history were collected through self-report and interview-based scales. Ancestry was identified using 292 genetic markers from the HapMap project. Using a regression analysis, we tested whether a history of migration, ethnicity and geographical ancestry were predictive of a history of suicide attempts.

**Results:** Our analysis failed to demonstrate a significant relationship between suicide history and migration, ethnicity or ancestry. However, ethnicity appears to be significantly associated with the number of psychiatric hospitalizations in our sample.

**Conclusion:** Although ethnicity and migration history are not predictive of previous suicide attempts, ethnicity may be an important aspect of psychiatric treatment, including access to mental health resources and frequency of hospitalizations.
3.2 Introduction

Suicide is a complex behaviour and a major public health concern, ranking among the leading causes of death in Canada (Statistics Canada, 2014). Despite being an avoidable outcome, suicide continues to contribute to the mortality of many individuals around the country. Due to the stigma and prejudice surrounding suicidal behaviour, many individuals fail to reach out or seek help when faced with these difficulties. Suicidal behaviour is generally defined along a continuum that includes suicidal ideation and suicide attempt, with the extreme end of the continuum being suicidal completion (Hor & Taylor, 2010). Unfortunately, due to the complex nature of suicide, it is difficult for clinicians to identify individuals at the highest risk for a suicide attempt.

Many individuals that attempt suicide often experience severe despair, frequently attributed to a psychiatric disorder. In fact, schizophrenia has been shown to increase an individual’s risk for suicide attempts (Saha et al., 2007). In Canada, there is arguably an increase in the rate of schizophrenia and higher prevalence and incidence rates compared to some other international populations (Dealberto, 2013). Schizophrenia is thought to arise due to a combination of environmental stressors and genetic factors with an average age of onset around late adolescence and early adulthood; leading to chronic impairment throughout the patient’s life (Holder & Wayhs, 2014). Though they vary among individuals, these symptoms typically include delusions, sensory hallucinations, flat or blunted emotions, and disorganized thoughts and speech (Fletcher & Frith, 2009; Phan & Kreys, 2011; Hovington & Lepage, 2012). As a major cause of disability, schizophrenia can have severe economic, health, and personal consequences on patients and those around them, especially without adequate treatment.
Individuals with schizophrenia generally experience a range of positive, negative, and cognitive symptoms (Phan & Kreys, 2011). Accordingly, schizophrenia is associated with a significantly increased risk for self-injurious and suicidal behaviors (Brown et al., 2000). In particular, suicide attempts are one of the largest contributors to the increased morbidity and mortality rates in schizophrenia (Hawton et al., 2005). It has been found that approximately between 20-40% of patients diagnosed with schizophrenia will attempt suicide during the duration of their illness (Pompili et al., 2007). Sadly, identifying individuals with schizophrenia at the highest risk for suicide attempt remains a major challenge for health practitioners.

Some studies have proposed that an increased incidence of schizophrenia is associated with being an ethnic minority and/or immigrant (Kirkbride et al., 2012; Fearon et al., 2006; Leao et al., 2006). It has been previously reported that immigrant populations have an increased risk for developing schizophrenia; suggesting that social factors play a vital role in the pathology of schizophrenia (Li et al., 2012; Cantor-Graae, 2007). Notably, belonging to a visible minority group also leads to a higher risk for developing a schizophrenia spectrum disorder. On the other hand, being a part of a neighbourhood with a high density of one’s own ethnic minority group can be a protective factor (Das-Munshi et al., 2012; Schofield et al., 2011; Kirkbride et al., 2007). Therefore, it has been proposed that rather than the stressors of immigration, it is being “the exception to the norm” during one’s lifetime that contributes to the increased risk for schizophrenia (van Os, 2012; Veling, 2013). Interestingly, a meta-analysis by Cantor-Graee and Selten (2005) found that the mean weighted relative risks of schizophrenia were 2.7 and 4.5 times higher for first and second generation immigrants, respectively. Cantor-Graee (2007) proposed that social defeat may be a plausible mechanism for explaining these elevated risks, whereby social defeat is defined as an experience of being in a subordinate or outsider position.
Not only do rates of psychosis differ between immigrants and native-born individuals, but there is also a difference in suicide attempt rates between these two groups in some countries (Malenfant, 2004). The experiences that come along with immigration are highly heterogeneous and not all immigrants face similar situations before or after migration (Bursztein Lipsicas & Henrik Mäkinen, 2010). Settling in a new country can be an extremely stressful situation for the individual, potentially causing severe psychological distress and triggering suicidal behaviours. Many individuals face drastic changes in their social roles and social status, and are potentially exposed to social marginalization, prejudice, and discrimination by the host population (Bhugra, 2005; Shoval et al., 2007). In comparison to the studies conducted so far on suicide and immigration, regrettably only a few studies have focused on attempted suicide among immigrants.

Previous research suggests that the immigrant population is at a higher risk of developing both schizophrenia and attempting suicide (Berg et al., 2014; Spallek et al., 2015). Immigrant and ethnic differences in suicide attempt behaviour can guide suicide prevention efforts because they are potential precursors to completed suicide. This is particularly important and relevant for ethnically diverse countries such as Canada, where studies have shown that immigrant and ethnically diverse populations show differences in suicide completion and suicide attempts (Dealberto 2013). Therefore, the aims of this study are to predict whether individuals with schizophrenia are more likely to have attempted suicide if they have immigrated to Canada during their lifetime, and whether differences in self-reported ethnicity or genetically-determined geographical ancestry are predictive of a history of suicide attempts. Given the current literature on the topic, it is hypothesized that individuals with schizophrenia who have migrated to Canada and are of an ethnically different background from the visible majority in Canada (European Caucasian) are more likely to have a history of attempted suicide.
3.3 Materials and Methods

3.3.1 Subjects and Assessments

We originally recruited 336 participants from the Centre for Addiction and Mental Health (CAMH) in Toronto, Canada between the ages of 18 to 75. All patients met the criteria for schizophrenia spectrum disorders based on the structured clinical interview for DSM-IV (SCID-I/P). Inclusion criteria included a diagnosis of schizophrenia or schizoaffective disorder while our exclusion criteria included evidence of intellectual disability and the presence of neurodegenerative disorders. In addition, those who experienced traumatic brain injury with a loss of consciousness and a history of major substance abuse prior to the onset of illness were excluded to ensure that the onset of the participant’s symptoms was not directly attributed to the intake of drugs or physical trauma. Written informed consent was obtained from all individual participants included in the study. This study was approved by the CAMH Research Ethics Board.

Assessments were conducted cross-sectionally using a structured interview and self-report questionnaires. The interview incorporated the Structured Clinical Interviews for DSM-IV (SCID-I/P) in order to diagnose participants, as well as to assess for additional psychiatric symptoms and comorbid diagnoses. In situations where a diagnosis could not be reliably defined, the individual was excluded from the analysis. Information regarding the participants’ ethnicity was also collected through self-report and interview questions. Ethnicity was defined by identifying the ethnic background of the participants’ four grandparents and through written self-report. Individuals were classified as either European Caucasian or Other. Cases in which there was ambiguity or discrepancy in the individuals ethnicity lead to them being excluded. The individuals in our sample classified as immigrants are defined as first-generation immigrants only.
During the clinical assessment, participants’ suicide attempt history was assessed using self-report and interview based measures including the Beck Suicide Ideation Scale (BSS) (Beck & Steer, 1991) and the Columbia Suicide Severity Rating Scale (C-SSRS) (Posner et al., 2011), respectively. According to the patients’ response on the BSS and C-SSRS, they were classified as suicide attempters (those who have attempted suicide at least once in their lifetime) and non-attempters (those who never attempted suicide in their lifetime). Individuals with an ambiguous or inconsistent suicide attempt history between the BSS and C-SSRS were excluded.

Our original sample size consisted of 336 individuals diagnosed with either schizophrenia or schizoaffective disorder; however, 276 participants were included after filtering criteria were applied. Participants not included in our analysis were those who had inconsistent, incomplete or inaccurate reports of previous suicide attempts and an ambiguous or unclear ethnic background. Furthermore, only patients who attempted suicide after immigration were included in the analysis. All those individuals who had attempted suicide prior to migrating to Canada were excluded.

3.3.2 Clinical Analysis on Suicide Attempt History

All data analyses were performed using R (v.3.1.3). Participants were assessed according to two groups: Suicide Attempters and Non-Attempters. The presence of at least one suicide attempt lifetime versus patients who never attempted suicide was used as the main grouping variable. Attempters and non-attempters were compared to test for significant differences regarding age, sex, ethnicity and immigrant status. A summary of the clinical variables can be found in Table 3-1. We used a logistic regression model to determine whether differences in immigration status (immigrants versus native-born individuals) and ethnicity (European Caucasian versus Others) were predictive of a history of previous suicidal attempts. Our tests of significance were done with a confidence interval of 95% with an alpha level of 0.05.
### Table 3-1. Clinical demographics of the study sample based on stratification of suicide attempters and non-attempters.

<table>
<thead>
<tr>
<th>Total (N=276)</th>
<th>Suicide Attempters (N=126)</th>
<th>Non-Attempters (N=150)</th>
<th>P-Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (M ± S)</td>
<td>39.74 ± 11.42</td>
<td>40.56 ± 12.42</td>
<td>0.435 &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Age of Onset (M ± S)</td>
<td>22.04 ± 6.71</td>
<td>22.52 ± 7.26</td>
<td>0.569 &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sex (Male/Female)</td>
<td>91/35</td>
<td>96/54</td>
<td>0.185 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Immigration Status (Immigrant/Native)</td>
<td>29/97</td>
<td>35/115</td>
<td>0.989 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hospitalizations (M ± S)</td>
<td>5.98 ± 6.95</td>
<td>4.29 ± 5.23</td>
<td>0.878 &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Comorbid Alcohol Abuse (%)</td>
<td>37%</td>
<td>28%</td>
<td>0.129 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Comorbid Drug Abuse (%)</td>
<td>31%</td>
<td>29%</td>
<td>0.778 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Marijuana Abuse (%)</td>
<td>25%</td>
<td>25%</td>
<td>0.482 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Duration of Illness (years) (M ± S)</td>
<td>17.73 ± 11.00</td>
<td>18.42 ± 11.70</td>
<td>0.281 &lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Self-Reported Ethnicity (European Caucasian/Other)</td>
<td>101/25</td>
<td>112/38</td>
<td>0.348 &lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

M ± S  = mean ± standard deviation;
<sup>a</sup> Chi-square test with Yates’ continuity correction
<sup>b</sup> Independent t-test

### 3.3.3 Genetic Analysis Subjects of Geographical Ancestry

Our sample consisted of 155 subjects genotyped using a customized Illumina Bead Chip and 121 subjects genotyped using the Illumina Omni-2.5 array. Genotype imputation was conducted for the 121 subjects using IMPUTE2 (v.2.3.1) and 1000 Genomes Phase 3 reference data to obtain a selection of 292 SNPs from an original panel of 384 SNPs represented in the HapMap Phase II project (Frazer et al., 2007). Three reference populations were used from the HapMap project: European Caucasians (North/Western Europeans from Utah [CEU]), East Asians (Han Chinese and Japanese individuals [CHB + JPT]), and Africans from Nigeria.
(Yoruba from Nigeria [YRI]). Our final analysis was done considering 276 individuals as 5 participants had incomplete genotype data and were excluded.

We determined ancestry with the Multidimensional Scaling (MDS) procedure using PLINK v1.07. The first two components were used to identify population stratification. Using R (v.3.1.3), we plotted the two components for all participants in our study sample relative to the individuals from the three major ethnic clusters from HapMap II as seen in Figure 3-1. Ancestry was determined using the 292 SNP markers which are ancestry-informative according to allele frequencies. Self-reported ethnicity in all our participants was confirmed using the MDS plot. Individuals who self-reported as European Caucasian and Other ethnicity were visually distinguishable. We excluded individuals with discordant self-reported ethnicity and geographical ancestry, in which self-reported ethnicity did not match geographical ancestry according to the MDS. Cut-off points for discordant ethnicity and ancestry are visualized. Parameter were drawn at $x = 0.05$ to distinguish European Caucasians from those of African ancestry, and $y = 0.05$ to separate European Caucasians from those of East Asian ancestry. We then performed a second regression analysis for suicide attempt history and ethnicity with a more precise ethnic background.
Figure 3-1. Multidimensional Scale Clustering according to Geographical Ancestry using 292 SNP markers from the HapMap Phase II project. Cut-off points for discordant ethnicity and ancestry are visualized at $x = 0.05$ to distinguish European Caucasians from those of African ancestry, and $y = 0.05$ to separate European Caucasians from those of East Asian ancestry.

To test whether geographical ancestry was predictive of suicide attempt history we used STRUCTURE (v.2.3.4) (Pritchard et al., 2000) to identify population substructure and estimate the geographic ancestry of the study participants (Figure 3-2). The three HapMap Phase II populations were also included. The output from STRUCTURE provides the relative ethnic proportion for each individual, namely the percentage that each individual was of CEU, CHB+JPT, and/or YRI ancestry. As our analysis considered ethnicity as European Caucasian or
Other, we considered the proportion that each individual was of European ancestry in our regression model to test whether it was predictive of a history of previous suicide attempts.

**Figure 3-2.** STRUCTURE analysis of self-reported ethnicity of study participants and reference population from HapMap. 1 = Study participants (n = 276); 2 (green) = Japanese and Han Chinese Asians (JPT + CHB), reference population (n = 90); 3 (red) = North/Western Europeans from Utah (CEU), reference population (n = 60); 4 (blue) = Yorubans from Nigeria (YRI), reference population (n = 60).

### 3.4 Results

Our sample of 276 participants consisted of 186 males and 90 females. There were 126 individuals with a history of attempted suicide, 97 of whom were native-born while 29 were immigrants to Canada. Furthermore, there were 150 individuals who had never attempted suicide, including 115 native-born and 35 immigrants. As we chose to categorize the ethnicity of our participants as either European Caucasian or Other, our sample consisted of 213 self-reported European Caucasians and 63 individuals of Other Ethnic backgrounds. After correcting ethnicity for geographical ancestry according to our MDS plot, our sample size included 247 individuals. Specifically, we excluded 29 discordant individuals who self-reported as non-European Caucasian yet clustered in the European ancestry sample (CEU).
In the analysis of the clinical variables, our logistic regression analysis demonstrated non-significant results in predicting suicide attempt history for all observed outcomes: age of symptom onset (OR = 0.99; 95% CI: 0.96 - 1.02; p = 0.570), gender (OR = 1.46; 95% CI: 0.879 - 2.46; p = 0.146), age (OR = 0.99; 95% CI: 0.97 - 1.01; p = 0.437). Furthermore, we found that an individual’s migration status did not significantly predict the number of hospitalizations (OR = 0.97; 95% CI: 0.91 - 1.02; p = 0.295). On the other hand, ethnicity was significantly predicted by the number of hospitalizations (OR = 1.11; 95% CI: 1.03 - 1.21; p = 0.011) such that being of a European Caucasian ethnic background predicted a higher number of hospitalizations.

In our analysis, we considered whether immigration status (being native-born to Canada or an immigrant) was predictive of a history of suicide attempt. We found that there was no significant association between migration status and suicide (OR = 0.98; 95% CI: 0.56-1.72; p = 0.950). Furthermore, when we considered whether ethnicity (European Caucasian versus Other ethnic backgrounds) was predictive of suicide attempt history, we found no significant association (OR = 1.37; 95% CI: 0.78 - 2.45; p = 0.280). Because immigrant status and ethnicity are often intertwined, we tested the interaction of these two variables as well. Unfortunately, similar non-significant results were found (OR = 0.90; 95% CI: 0.41 -1.93; p = 0.787). When we performed a second regression analysis considering self-reported ethnicity, in which individuals with discordant ethnicities and geographical ancestries were excluded according to our MDS plot, we similarly found no significant relationship (OR = 1.89; 95% CI: 0.89 - 4.20 ; p = 0.105) between ethnicity and suicide attempt history. Additionally, we continued to see that the number of hospitalizations were significantly predicted by ethnicity according to self-report and confirmed by geographical ancestry (OR = 1.05; 95% CI: 1.01 - 1.10; p = 0.040).
Finally, we wanted to test whether geographical ancestry was able to reliably predict suicide attempt history. Figure 3-2 illustrates the STRUCTURE analysis of study participants (Study participants, n = 276; reference population of CEU, n = 60; reference population of JPT + CHB, n = 90; and reference population of YRI, n = 60). Ethnicity according to our STRUCTURE analysis was not significant with a history of previous suicide attempts in our schizophrenia population (OR = 0.18, 95% CI: -0.025 – 0.377, p = 0.088).

Unfortunately, given the small sample size, our study was not sufficiently powered to detect a statistically significant effect. Our logistic regression analysis of immigrant status and ethnicity had less than 80% chance of identifying a statistically significant effect at the observed magnitude given our post hoc power analysis. An effect size of 1.46 would be required to reach approximately 80% power. Statistical power analyses were done using G*Power 3 (Faul et al., 2009).

3.5 Discussion

In the current study, we aimed to observe whether migration status, ethnicity and geographical ancestry significantly predict a history of suicide attempts. From our analysis we found no significant association between self-reported ethnicity (European Caucasian versus Other ethnicity) or migration status (immigrant versus native-born) with previous suicide attempt history. After correcting self-reported ethnicity according to our MDS plot, there continued to be no significant association with suicide attempt history. Interestingly, we found that self-reported ethnicity significantly predicted the number of hospitalizations, such that those of a self-reported European Caucasian ethnicity had a higher number of hospital admissions compared to non-self-identified European Caucasians. Finally, we found that ancestry according to our STRUCTURE analysis was similarly not predictive of past suicide attempts.
Here we report that self-reported European Caucasian ethnicity is predictive of higher psychiatric hospitalizations. It is possible that the lower usage rates of mental health services by non-Caucasian individuals is a result of cultural and language barriers (Lindert et al., 2008). This may be especially true for the immigrant populations that face greater barriers to healthcare access, which may include low socioeconomic status; however, we found no significant relationship in this regard. In contrast, in a systematic review by Bhui et al. (2003), it was reported that, overall, individuals of African ancestry had the highest rates of inpatient admission. On the other hand, Cole et al. (1995) demonstrated similar results to our own in which Caucasians had higher rates of hospitalizations. A controversial issue relating to hospitalizations is that there have been differences in psychiatric admission rates in the United Kingdom in the minority population compared to the European Caucasian majority (Morgan et al., 2005; Singh et al., 2007). In culturally diverse populations it is important to address the issue of ethnicity and its role in mediating the efficacy of mental health services.

It is very likely that ethnicity has a strong influence on the direction, quality, and frequency of mental health services. Ethnicity typically dictates a unique social group identified by language; religion; and birth place (Anderson et al., 2014). As a result, there are unique differences between different ethnic groups that mediate decisions on preferences for seeking help for psychiatric concerns and the resources that are available to them (Cauce et al., 2002; Commander et al., 1999). Additionally, interactions with health care providers may be influenced by perceived or actual differences in ethnic groups (McGovern et al., 1994). Many studies have previously reported differences between ethnic groups in relation to access to mental health services; however, it is important to be cognizant of other factors such as socioeconomic status, discrimination, and language barriers that also influence health care access (Jarvis 2007; Snowden & Yamada, 2005).
When interpreting studies that aim to identify a relationship with ethnicity such as the current study, it is important to be aware of how ethnicity is defined. Here, we initially identified ethnicity according to participants’ self-report when performing our analyses. In this case, self-reported ethnicity may more likely be dependent upon culture. However, we also took further steps to identify ethnicity by incorporating geographical ancestry as a biological indicator of ethnicity by using ancestry-informative genetic markers. By looking at whether self-reported ethnicity matched that of biological geographical ancestry in our sample we found 29 individuals who self-reported as being non-European Caucasian but clustered in the European ancestry sample. Upon closer inspection, a majority of these individuals were of East Indian ancestry. Therefore, it is evident that the number of genetic markers used in this analysis was not sufficient to clearly separate populations with more closely shared ancestry. While ethnicity according to both self-report and biological ancestry were not significantly predictive of suicide attempt history, this procedure offers a novel and more robust measure of defining ethnicity for future studies of this nature. Future studies should, however, incorporate a large number of ancestry-informative markers to reliably cluster individuals according to geographical ancestry.

Lastly, our analysis failed to demonstrate predictive value for previous suicide attempt history when considering patients’ migration status. The act of immigrating has been thoroughly studied as a risk factor for the development of schizophrenia; however, it appears in our sample that this stressor may not further predict suicide attempts in schizophrenia. Generally it has been reported that immigrants from nations with low suicide rates tend to maintain these in their new country (Voracek et al., 2009). Similarly, immigrants from nations with high suicide rates maintain similar rates following migration. However, other studies have suggested that immigrants generally show higher suicide rates compared to their countries of origin (Hjern & Allebeck, 2002). To add to the complexity of migration and suicide, some studies show that
immigrant suicide rates are lower, or converge over time with those of the host country (Singh & Hiatt, 2006). One prevailing theory is that some of the immigrants who exhibit suicidal behaviour in the new country had suicidal tendencies in their country of origin, which may in turn be related to genetic susceptibility manifesting itself at times of severe distress (Voracek et al., 2009).

Given that our study focused solely on first-generation immigrants, it is necessary for future studies to consider second-generation immigrants and those who migrate due to political or crisis situations in their country of origin. Since each immigrant group faces different obstacles and stressors it is important to consider each group separately. Furthermore, there is a need for international multi-site prospective studies to better identify the predictors for suicide attempt by collecting a larger immigrant population and following their psychiatric treatment outcomes.

Unfortunately, due to the nature of this study, it is difficult to disentangle the effects that ethnicity and immigration status have on suicide in schizophrenia. Given that immigrants also belong to many visible minority groups, it is unclear which categorization has stronger predictive value for suicide attempts. Furthermore, aggregating immigrant groups as a singular entity can mask significant variations between different immigrant populations. The type and process of migration of course has a role, particularly for immigrants who are refugees or asylum seekers (Spallek et al., 2015). As such, comparisons of immigrant suicide rates with native populations have produced divergent results. Of course, it is important to remain cognizant of the fact that immigration and ethnicity comprise a multitude of additional factors that were not captured in this study. It is believed that religious beliefs, language, and living conditions all play a significant role in the perceived stress off ethnic minority and immigrant groups. Therefore,
future analyses should strive to include these important variables in order to identify any meaningful associations between immigration and ethnicity in predicting the risk for suicide attempts.

3.6 Conclusion

Our study demonstrates that immigrant status and ethnicity may not be predictive of previous suicide attempts in schizophrenia. However, ethnicity has been shown to influence the number of psychiatric hospitalizations and perhaps access to mental health services in the Canadian population. We also present a novel approach of implementing the principal component analysis with ancestry-informative SNP markers to obtain a robust classification of self-reported ethnicity which is supported by geographical ancestry. By using ancestry-informative markers, we were able to accurately identify the ethnic origin of our sample and determine whether ancestry defined by biological terms is able to predict suicidal behaviour in schizophrenia patients. While we failed to demonstrate any significant findings, immigrant studies are uniquely informative for investigating the genetic and environmental influences on suicidal behaviour in schizophrenia due to immigrants carrying different ancestry-informative markers and experiencing a potentially stressful life event.
Chapter 4

4 Classification of Suicide Attempt History in Schizophrenia using Sociocultural and Clinical Features

4.1 Abstract

Objective: Suicide is a major concern for those afflicted by schizophrenia. Identifying patients at the highest risk for future suicide attempts remains a complex problem for psychiatric intervention. Machine learning models allow for the integration of many risk factors in order to build an algorithm that predicts which patients are likely to attempt suicide. Currently it is unclear how to integrate previously identified risk factors into a clinically relevant predictive tool to estimate the probability of a patient with schizophrenia for attempting suicide.

Methods: We conducted a cross-sectional assessment on a sample of 345 participants diagnosed with schizophrenia spectrum disorders. Suicide attempters and non-attempters were clearly identified using the Columbia Suicide Severity Rating Scale (C-SSRS) and the Beck Suicide Ideation Scale (BSS). We developed two classification algorithms using a regularized regression and random forest model with sociocultural and clinical variables as features to train the models.

Results: Both classification models performed similarly in identifying suicide attempters and non-attempters. Our regularized logistic regression model demonstrated an accuracy of 66% and an area under the curve (AUC) of 0.71, while the random forest model demonstrated 65% accuracy and an AUC of 0.67.

Conclusion: Machine learning algorithms offer a relatively successful method for incorporating many clinical features to predict individuals at risk for future suicide attempts. Increased
performance of these models using clinically relevant variables offers the potential to facilitate early treatment and intervention to prevent future suicide attempts.

4.2 Introduction

Schizophrenia is a mental disorder which is associated with higher incidences of suicidal behaviour (Palmer et al., 2005, Fialko et al., 2006) which remains the leading cause of death among psychiatric patients. Clinicians attempting to treat patients with schizophrenia with no known previous history of suicide attempts must determine the likelihood that their patient will attempt suicide in the future, including during the course of their treatment. Being able to accurately predict or identify those at risk for future suicide attempts is a major challenge in psychiatry and in particular among schizophrenia patients. Many previous studies have attempted to identify clinical variables that may help predict the risk for suicide in schizophrenia. However, such efforts have largely been unsuccessful, producing many false positives or inconsistent results between studies (Pompili et al., 2007). Suicide is a leading cause of mortality among this diagnostic group but preventable with clinical interventions (Mann et al., 2005). Creating a statistical framework that successfully incorporates the role of potential clinical and sociocultural risk factors may allow for an accurate prediction of suicide attempt risk for an individual patient. Being able to predict suicide risk at the individual level will undoubtedly allow for improved efforts in mitigating future suicide attempts among high risk patients in this population (Fenton, 2000).

Suicide is a highly complex behaviour and a major public health concern. There appears to be a significant increase in the risk for suicide among those with schizophrenia as compared to the general population (Brown et al., 2000). Previous epidemiological studies have conservatively suggested a 4.9% risk for suicide among this population (Palmer et al., 2005) and
approximately 20-40% attempting suicide during their lifetime (Pompili et al., 2007). While many patients may attempt suicide due to psychosocial stressors associated with the disorder, there are also a significant number of patients with the same stressors who do not attempt suicide. Many risk factors have been proposed to explain why some individuals may be driven to attempt suicide (Popovic et al., 2014) but these mainly delineate the differences between suicide attempters and non-attempters without adding predictive power. In order to improve the predictive accuracy of these risk factors, it is imperative that they be integrated to establish a predictive model that may be able to estimate the likelihood of an individual patient with schizophrenia to attempt suicide.

The prediction of suicide attempters and non-attempters in schizophrenia is a complex classification problem that requires the simultaneous inclusion of many potential risk factors (Ribeiro et al., 2016a). Classification is defined as the process of assigning individuals to one or more classes – in this case, suicide attempters or non-attempters. Machine learning algorithms in psychiatric research are emerging as a valuable tool to incorporate many predictor variables and improve classification capabilities among unseen observations (i.e. patients) (Passos et al., 2016). Algorithms that include a small number of rules or operations, such as in general linear models, may likely solve simple classification problems in psychiatry. However, complex classification problems, such as suicide attempt status, require algorithms that model complex relationships among a large number of factors (Ribeiro et al., 2016b; Bentley et al., 2016). Machine learning models are designed to consider these complex relationships so as to determine an optimal classification algorithm. In fact, recent studies have demonstrated some success in utilizing these models in suicide research among U.S. soldiers (Kessler et al., 2015).
Many individual risk factors have been put forward that may contribute to suicide risk in schizophrenia. To develop our algorithm, we considered sociocultural and clinical risk factors suggested in the literature. Feature (variable) selection for our model was based on *a priori* evidence from the scientific literature of sociocultural risk factors for suicide attempts associated with demographic, illness-related, sociocultural, comorbid, and clinical scale variables. In the current study we implemented two machine learning algorithms: a least absolute shrinkage and selection operator (LASSO) regularization penalty with a general binary logistic model (Tibshirani, 2011) and Random Forests (RF) (Breiman, 2001). Our main objective was to establish a clinically useful predictive model able to accurately identify suicide attempters and non-attempters among our sample of schizophrenia patients.

4.3 Materials and Methods

4.3.1 Study Sample

We recruited a final sample of 345 participants from the Centre for Addiction and Mental Health (CAMH). Participant inclusion criteria included those between the ages of 18 to 75 and who met the criteria for schizophrenia spectrum disorders based on the structured clinical interview for DSM-IV (SCID-I/P). Evidence of intellectual disability, a history of brain injury trauma with a loss of consciousness, and previous major substance abuse prior to the onset of the psychiatric illness were considered as exclusion criteria. Written informed consent was obtained for participation in the study as well as for the release of participants’ medical history in order to verify clinical data or obtain missing information. This study was approved by the CAMH Research Ethics Board.
4.3.2 Clinical Assessments

We conducted cross-sectional assessments using structured interview and self-report questionnaires. The interview incorporated the Structured Clinical Interviews for DSM-IV (SCID-I/P) to confirm the participants’ primary diagnosis and assess for additional psychiatric symptoms and comorbid diagnoses (i.e. lifetime alcohol abuse/dependence and drug abuse/dependence). In cases where a reliable diagnosis could not be clearly defined, the individual was excluded from the analysis. Participants’ suicidal history was assessed using the Beck Suicide Ideation Scale (BSS) (Beck & Steer 1991) and the Columbia Suicide Severity Rating Scale (C-SSRS) (Posner et al., 2011). According to the patients’ response on the BSS and C-SSRS, they were classified as suicide attempters (those who have attempted suicide at least once in their lifetime) and non-attempters (those who never attempted suicide in their lifetime). Individuals with a contradictory suicide attempt history between self-report and interview-based scales were confirmed using the medical charts and unresolved cases were excluded. Information regarding the participants’ ethnicity was collected through self-report and interview questions. Individuals were classified as either European Caucasian or Other. We also administered the Childhood Trauma Questionnaire (CTQ) and NEO Five Factor Inventory (NEO-FFI) scale at the time of the assessment. Further clinical information was collected regarding familial history of suicidal behaviour, psychosis, and mood disorders. Information on the participants’ primary language and religious affiliation were collected.

4.3.3 Predictor Variable Selection

We included 26 predictor variables in our model based on a priori evidence for an increased risk of suicide attempt in schizophrenia. The variables selected were associated with immigration and ethnicity, including whether or not they were born in Canada (Berg et al., 2014; Spallek et al., 2015); their age at immigration (Malenfant, 2004); first language (English or
Other) and religion (Religious Affiliation or No Religion) (Rosmarin et al., 2013). Furthermore, important clinical risk factors putatively associated with suicide risk in schizophrenia were included such as age (Hor & Taylor 2010); sex (Hawton et al., 2005); duration of illness (Carlborg et al., 2010); number of hospitalizations (Togay et al., 2015); family history of mood disorder (Sanchez-Gistau et al., 2013), psychosis (Björkenstam et al., 2014), and suicide (Qin et al., 2002; Trémeau et al., 2001), and lifetime history of alcohol and drug abuse/dependence (Togay et al., 2015). Scores from clinical measures previously associated with suicide attempt were also incorporated including NEO-FFI (Pillmann et al., 2003) and CTQ (Hassan, Stuart, & De Luca, 2016). Since the predictor variables have different value ranges in which some are discrete variables and others are continuous variables they were scaled according to z-score normalization using Python. The selected predictor variables were used in the machine learning algorithms to predict the classification output labels (suicide attempter or non-attempter).

Demographic and clinical data with predictor variable characteristics are summarized in Table 4-1. We performed a univariate logistic regression analysis for all predictor variables with suicide attempt status as the outcome. All statistical analyses were carried out in R v. 3.3.0.

### 4.3.4 Classification Algorithms

The clinical sample was split according to a history of suicide attempt (suicide attempters = 1; suicide non-attempters = 0). Our goal was to correctly classify individuals in our sample based on this binary classification. We utilized two machine learning algorithms for classification (LASSO and RF) with Python 2.7, using the scikit-learn package, which were run on the Specialized Computing Cluster (SCC) at CAMH. The LASSO technique implements a standard logistic regression formula but adds a ‘penalization’ procedure by assigning some predictor variable weights as zero. This process is referred to as regularization. Regularization improves the predictive accuracy of the binary logistic regression model and allows for improved...
generalizability to different datasets. Penalization leads to some of the parameter estimates being exactly zero. The RF model works by creating a multitude of decision trees and providing an output class. RF produces a collection of randomly generated decision tree predictors. Each branching node is equivalent to one of the input variables. RF selects a subset of input variables to construct a tree, whereby each tree decides the most likely class for the given inputs and picks the most frequent class among all the trees. The overall goal of both algorithms is to create a model that predicts the classification label based on many input variables.

4.3.5 Model Testing and Cross-Validation

Overfitting is an issue when a classifier fits the training data too closely such that random error and noise are described in the model. This makes it difficult then for the model to generalize trends to new data. Additional techniques, such as cross-validation and regularization, can help avoid overfitting. To test our models, it is common practice when performing a (supervised) machine learning technique to hold out part of the available data as a ‘test set’. In a given dataset there may be a class imbalance in which the participants in one class (suicide non-attempters) are more frequent than participants in the other class (suicide attempters) (Dubey et al., 2014). When splitting the data into ‘training’ and ‘testing’ dataset, it is ideal for the class distribution to be preserved. In our splitting of training and testing data we incorporated a Stratified K-Fold cross-validation procedure. The data was first partitioned into $k$ equally sized folds. The folds were made by preserving the ratio for suicide attempters and non-attempters. Here, we assigned $k = 10$, whereby 10 iterations of training and validation were done, such that within each iteration, each fold is used once as the test set while the remaining $k-1$ folds are used for training. Prior to being split into $k$ folds, the data was stratified by rearranging the data so as to ensure that each fold is a good representative of the entire dataset. In our current classification problem, there are an unequal number of attempters and non-
attempters. Therefore, stratification was done to maintain a balance between these two labels in each fold, as much as possible. Our performance summary statistics were calculated as an average across all folds. Average accuracy and area under the curve (AUC) scores with their standard errors (SE) are generated using values averaged across all folds. A summary of the procedure is illustrated in Figure 4-1.

Figure 4-1. Flow chart demonstrating the process by which our machine learning algorithms trained and tested our clinical variables in predicting suicide attempters. Stratified K-fold Cross validation was used to maintain class distributions across all folds; T= Test Data.
Table 4-1. Summary statistics for clinical, demographic, and sociocultural predictor variables. Predictor variable type is indicated as either categorical or continuous.

<table>
<thead>
<tr>
<th>Total (N= 345 )</th>
<th>Suicide Non - Attempter (N = 214)</th>
<th>Suicide Attempter (N = 131)</th>
<th>P-value</th>
<th>Predictor Variable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) (M ± S)</td>
<td>38.89 ± 12.92</td>
<td>44.51 ± 12.62</td>
<td>1.35 x 10^{-4}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>Sex (Male/Female)</td>
<td>157/57</td>
<td>84/47</td>
<td>0.070</td>
<td>Categorical variable: 1- male; 0 – female</td>
</tr>
<tr>
<td>Duration of Illness (M ± S)</td>
<td>15.91 ± 11.39</td>
<td>22.09 ± 12.52</td>
<td>7.97 x 10^{-6}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>Number of Hospitalizations (N&gt;5)</td>
<td>50 (23%)</td>
<td>61 (47%)</td>
<td>1.05 x10^{-3}</td>
<td>Categorical variable: 1 – (&gt; 5); 0 – (&lt;5)</td>
</tr>
<tr>
<td>Immigrated before age 9</td>
<td>22 (10%)</td>
<td>19 (15%)</td>
<td>0.241</td>
<td>Categorical variable: 1- immigrated before 9; 0 – N/A</td>
</tr>
<tr>
<td>Immigrated between ages 10-18</td>
<td>31 (15%)</td>
<td>16 (12%)</td>
<td>0.551</td>
<td>Categorical variable: 1- immigrated between 10-18; 0 – N/A</td>
</tr>
<tr>
<td>Immigrated after age 18</td>
<td>14 (7%)</td>
<td>11 (8%)</td>
<td>0.520</td>
<td>Categorical variable: 1- immigrated after 18; 0 – N/A</td>
</tr>
<tr>
<td>Self-Reported Ethnicity</td>
<td>108 (55%)</td>
<td>77 (59%)</td>
<td>0.134</td>
<td>Categorical variable: 1- European Caucasian; 0 – Other</td>
</tr>
<tr>
<td>Religious Identity</td>
<td>155 (72%)</td>
<td>103 (79%)</td>
<td>0.199</td>
<td>Categorical variable: 1- religious identity; 0 – no religion</td>
</tr>
<tr>
<td>English as Primary Language</td>
<td>183 (86%)</td>
<td>111 (85%)</td>
<td>0.843</td>
<td>Categorical variable: 1- English; 0 – other language</td>
</tr>
<tr>
<td>Family History of Psychosis</td>
<td>57 (27%)</td>
<td>51 (39%)</td>
<td>0.174</td>
<td>Categorical variable: 1- family history; 0 – N/A</td>
</tr>
<tr>
<td>Family History of Mood Disorders</td>
<td>76 (36%)</td>
<td>65 (50%)</td>
<td>0.010</td>
<td>Categorical variable: 1- family history; 0 – N/A</td>
</tr>
<tr>
<td>Family History of Suicidal Behavior</td>
<td>34 (16%)</td>
<td>47 (36%)</td>
<td>3.19 x 10^{-3}</td>
<td>Categorical variable: 1- family history; 0 – N/A</td>
</tr>
<tr>
<td>NEO Neuroticism (M ± S)</td>
<td>33.47 ± 8.36</td>
<td>36.75 ± 8.61</td>
<td>7.14 x 10^{-4}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>NEO Extraversion (M ± S)</td>
<td>37.19 ± 6.60</td>
<td>37.86 ± 7.67</td>
<td>0.672</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>NEO Openness (M ± S)</td>
<td>39.11 ± 6.66</td>
<td>39.24 ± 6.46</td>
<td>0.856</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>NEO Agreeableness (M ± S)</td>
<td>40.34 ± 5.76</td>
<td>40.40 ± 5.20</td>
<td>0.958</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>NEO Conscientiousness (M ± S)</td>
<td>42.45 ± 7.55</td>
<td>41.41 ± 7.82</td>
<td>0.221</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Physical Abuse (M ± S)</td>
<td>7.51 ± 3.69</td>
<td>9.66 ± 5.25</td>
<td>3.39 x 10^{-3}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Emotional Abuse (M ± S)</td>
<td>9.68 ± 4.58</td>
<td>12.90 ± 5.99</td>
<td>2.03 x 10^{-7}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Sexual Abuse (M ± S)</td>
<td>7.29 ± 4.09</td>
<td>8.85 ± 5.39</td>
<td>3.23 x 10^{-4}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Physical Neglect (M ± S)</td>
<td>8.48 ± 3.49</td>
<td>9.56 ± 4.29</td>
<td>0.013</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Emotional Neglect (M ± S)</td>
<td>11.07 ± 4.81</td>
<td>13.24 ± 5.76</td>
<td>2.54 x 10^{-4}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>CTQ Total Score (M ± S)</td>
<td>38.35 ± 4.74</td>
<td>47.09 ± 19.20</td>
<td>9.91 x 10^{-6}</td>
<td>Continuous variable</td>
</tr>
<tr>
<td>Lifetime Alcohol Abuse/ Dependence</td>
<td>63 (29%)</td>
<td>37 (28%)</td>
<td>0.812</td>
<td>Categorical variable: 1- alcoholism; 0 – N/A</td>
</tr>
<tr>
<td>Lifetime Drug Abuse/Dependence</td>
<td>62 (29%)</td>
<td>52 (40%)</td>
<td>0.041</td>
<td>Categorical variable: 1- alcoholism; 0 – N/A</td>
</tr>
</tbody>
</table>

M ± S = mean ± standard deviation
Model comparisons were done using Receiver Operating Characteristic (ROC) curves and their corresponding AUC score. In order to assess the validity of our models in correctly identifying individuals as suicide attempters or non-attempters we calculated specific model statistics, including prediction accuracy, sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV). Sensitivity is referred to as the true positive rate which measures the proportion of positives (suicide attempters) that were correctly identified. Specificity (true negative rate) is a measure of the proportion of negatives (suicide non-attempters) that were correctly identified. Predictive accuracy was calculated as the sum of true positives and true negatives, divided by the total sample. PPV (%) was calculated as the proportion of positive results that are true positives and NPV (%) was calculated as the proportion of negative results that are true negatives.

4.4 Results

Our final sample consisted of 345 patients with schizophrenia spectrum disorders who were used in the development of our models. All individuals included in the model had no missing data for the included predictor variables. Any missing data was retrieved through medical chart review when possible. Suicide attempters and non-attempters were significantly predicted by age (OR = 1.034; 95% CI: 1.017 - 1.052; p = 1.35 x 10^{-4}), duration of illness (OR = 1.043; 95% CI: 1.024 - 1.063; p = 7.97 x 10^{-6}), and number of hospitalizations (OR = 2.858; 95% CI: 1.797 - 4.578; p = 1.05 x 10^{-5}). Individuals who attempted suicide were older, with more hospitalizations of 5 or more, and have lived with their illness for longer. Furthermore, we found that a higher frequency of family history for mood disorders (OR = 1.788; 95% CI: 1.150 - 2.789; p = 0.010) and suicidal behaviour (OR = 2.962; 95% CI: 1.782 - 4.972; p = 3.19 x 10^{-5}) was significantly associated with suicide attempt history. Among the clinical scales we
administered, suicide attempt history was associated with NEO Neuroticism (OR = 1.047; 95% CI: 1.020 - 1.076; p = 7.14 x 10^-4) and CTQ items for Physical Abuse (OR = 1.114; 95% CI: 1.060 - 1.175; p = 3.39 x 10^-5), Emotional Abuse (OR = 1.121; 95% CI: 1.075 - 1.171; p = 2.03 x 10^-7), Sexual Abuse (OR = 1.072; 95% CI: 1.024 - 1.125; p = 3.23 x 10^-3), Physical Neglect (OR = 1.075; 95% CI: 1.016 - 1.139; p = 0.013), Emotional Neglect (OR = 1.082; 95% CI: 1.038 - 1.130; p = 2.54 x 10^-4), and total CTQ score (OR = 1.031; 95% CI: 1.018 - 1.046; p = 9.91 x 10^-6). Lastly, a history of drug abuse/dependence was also significantly associated with suicide attempt history (OR = 1.614; 95% CI: 1.020 - 2.553; p = 0.041).

As the goal of our analysis is to correctly identify individual suicide attempters from our sample of schizophrenia patients, it is imperative that we obtain the lowest false negative rate possible. For studies of this nature, it is preferable to avoid falsely dismissing a patient as a non-attempter for fear that this patient may attempt suicide in the future. Namely, false positives are preferable to false negatives. Therefore, the best model would aim to have a high degree of sensitivity as its main priority. Given the two classifier models, we reported below the summary statistics which compare the efficacy of both models in predicting suicide risk.

The LASSO model was able to correctly identify suicide attempters and non-attempters with a cross-validated prediction mean accuracy of 0.66 ± 0.02. The corresponding ROC curve (Figure 4-2) and confusion matrix (Figure 4-3) for all stratified folds were used to calculate a cross-validated sensitivity and specificity. The algorithm was able to generate an average sensitivity of 76% and average specificity of 56%. By creating ROC curves for all ten stratified folds, we generated an average AUC of 0.71 ± 0.02. We also calculated an average PPV of 66% and a NPV of 70% from our model.
**Figure 4-2.** ROC curve for LASSO regression model across all stratified folds in cross-validation. A final AUC score is given as the average of all AUC values across all 10 folds.
The RF model was able to correctly identify suicide attempters and non-attempters with a cross-validated prediction mean accuracy of $0.65 \pm 0.02$. The corresponding confusion matrix (Figure 4-4) and ROC curve (Figure 4-5) for all stratified folds were used to calculate a cross-validated sensitivity and specificity. The algorithm was able to generate an average sensitivity of 69% and an average specificity of 55%. Using 10-fold ROC curves we generated an average AUC of $0.67 \pm 0.03$. We also calculated a PPV of 61% and a NPV of 63% from our model. A comparison of the performance of both LASSO and RF algorithms is given in Table 4-2.

Within the LASSO regression model, the top five predictor variables averaged across all folds for the model were duration of illness, number of hospitalizations (>5), CTQ emotional abuse item, lifetime drug abuse/dependence, and the CTQ physical abuse item (Figure 4-6).
These variables were given positive weights in order to classify suicide attempters in our sample. Sex and lifetime alcohol abuse/dependence were consistently assigned a negative weight across all folds. Predictor variables that increase the likelihood of classifying a patient as a suicide attempter are given positive weights, in contrast to predictor variables with negative weights that indicate a decrease in their value increases the probability of correctly identifying a suicide attempter. Similarly, the RF model demonstrated consistent predictor variables across all folds. Unlike the LASSO model which sets certain predictor variable weights to zero in the model, RF uses all predictor variables to classify observations. The top five predictor variables averaged across all folds for the RF model were age, CTQ emotional abuse, duration of illness, CTQ total score, and NEO Neuroticism (Figure 4-7).

![Image of confusion matrix]

**Figure 4-4.** Mean confusion matrix describing performance of the RF model across all stratified folds from the cross-validation.
Figure 4-5. ROC curve for Random Forest model across all stratified folds in cross-validation. A final AUC score is given as the average of all AUC values across all 10 folds.
Figure 4-6. Visualization of each predictor variable and their average weight in the LASSO regression model for all 10 cross-validation folds. Each variable’s weight indicates the level of importance in predicting the classification phenotype following regularization. Predictor variables demonstrate a positive, negative, or zero weight due to penalization.
Figure 4-7. Visualization of each predictor variable and their average weight in the RF model for all 10 cross-validation folds. Each variable’s weight indicates the level of importance in predicting the classification phenotype following regularization.
Table 4-2. Comparison of algorithm performance in classifying individual patients as either suicide attempters or non-attempters. Values are provided as means and standard errors.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>LASSO</th>
<th>Random Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area under curve (AUC)</td>
<td>0.71 ± 0.02</td>
<td>0.67 ± 0.03</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.66 ± 0.02</td>
<td>0.65 ± 0.02</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.76 ± 0.02</td>
<td>0.69 ± 0.02</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.56 ± 0.02</td>
<td>0.55 ± 0.02</td>
</tr>
<tr>
<td>PPV</td>
<td>0.66 ± 0.02</td>
<td>0.61 ± 0.03</td>
</tr>
<tr>
<td>NPV</td>
<td>0.70 ± 0.02</td>
<td>0.63 ± 0.03</td>
</tr>
</tbody>
</table>

4.5 Discussion

Many studies have attempted to identify clinically relevant risk factors to preemptively identify individuals with schizophrenia who are most likely to attempt suicide. Unfortunately, the strength of these risk factors varies between studies and largely produces false positive effects. Although many risk factors have been put forward, it has remained an elusive goal to synthesize these into a clinically relevant tool to predict patients who are at the greatest risk for suicide attempt, at the individual level. By using machine learning algorithms, this is the first study, to our knowledge, to bring together many potential risk factors to identify the risk for schizophrenia patients attempting suicide. Our algorithms demonstrated a predictive accuracy of 65-66% in correctly identifying suicide attempters and non-attempters, which is greater than what would have been determined by chance alone. Both the LASSO and RF model performed equally well, in terms of the aforementioned prediction accuracy with AUC scores of 0.71 and 0.67, respectively. Our model statistics were validated using a stratified k-fold cross-validation approach. This prevented an inherent bias in our training and testing data to ensure that both models were able to train on data which maintained the original distribution of suicide attempters and non-attempters. Our machine learning algorithms were ultimately able to determine the
probability of distinguishing an individual as a suicide attempter or non-attempter with confidence using the predictor variables we included.

The predictor variables included in our model were based on *a priori* knowledge of risk factors for suicide attempt in schizophrenia from the scientific literature. Current classification attempts in psychiatry rely on the observation of individual symptoms to reliably make a clinical diagnosis. This is particularly difficult for many psychiatric phenotypes, especially suicide attempting. Suicidal behaviour ranges on a continuum and vary in their lethality and risk. Here, we focused on classifying individuals with schizophrenia spectrum disorders as either suicide attempters or non-attempters. Particularly in schizophrenia, suicide attempts may occur without any obvious warnings or clear signs; making it difficult to predict (Togay et al., 2015). Unique to schizophrenia spectrum disorders, patients may be driven to attempting suicide due to delusional thoughts or hallucinations. Therefore, suicide attempting is difficult to categorize under these circumstances. Poor objective methods of determining intentionality, some regard suicide attempts as “parasuicide” or “deliberate” or “non-fatal” self-injury instead, where there is believed to be an intention to die at least greater than zero. Furthermore, suicide attempts vary widely in their lethality and method choices. Classifying all suicide attempters as a single diagnostic label may therefore mask the subtleties in risk factors that lead to different suicidal behaviours. Because the diagnostic reliability of a suicide attempt as a phenotype may lack some reliability and under-represents inter-individual differences, using suicide attempt status as a classification label may prove to be ineffective in developing robust machine learning predictive models (Freedman et al., 2013). Having a clear, distinctive classification label is one way in which these models will improve in their reliability and generalization to a larger population, which may require an improved system for subtyping behavior we currently term loosely as suicide attempts.
Although our study is arguably the first to attempt to utilize machine learning algorithms to predict suicide attempters from non-attempters in a sample of schizophrenia patients, other studies have attempted to do the same in other diagnostic domains. In one study, a signature for suicide attempts in mood disorders was derived using machine learning models (Passos et al., 2015). The authors found in their most predictive model, using a relevance vector machine (RVM) algorithm, an AUC of 0.77 with 72% accuracy. Compared to our LASSO model which was our most predictive with an AUC of 0.71 and 66% accuracy our algorithms performed quite respectably for the given phenotype. Despite using a much smaller sample size, different machine learning techniques, different predictor variables, and investigating suicide attempt in schizophrenia as opposed to mood disorders, our results do not substantially differ. This leaves the open question as to the effectiveness of these models in predicting suicide attempt risk across many diagnostic labels.

Using clinical and demographic predictor variables alone may not be sufficient in predicting the risk for suicide. Future studies may benefit from incorporating a biological hypothesis for suicide attempt in which candidate genes (de Medeiros Alves et al., 2015) or imaging data (Lee et al., 2016; Minzenberg et al., 2014) may be included in these models. In studies that have attempted to identify risk factors associated with suicide in schizophrenia, many of the same risk clinical factors have been repeatedly suggested (Togay et al., 2015). However, no studies to our knowledge, have attempted to perform a large-scale analysis of risk factors for suicide attempt in schizophrenia using clinical, genetic, and imaging data to define a clear risk for suicide. Future studies of this kind may therefore benefit from the incorporation of risk factors clearly identified by these multimodal studies that combine data across all dimensions.

To our knowledge, as the first study to use machine learning algorithms to classify suicide attempters and non-attempters in schizophrenia, our models are only reasonably predictive.
Notably, compared to other classification studies for suicide (Passos et al., 2016; Kessler et al., 2015; Guan et al., 2015) our models perform nearly as effectively. The limitations of our study, which may have prevented higher AUC and accuracy scores, include our relatively small sample size and limited number of predictor variables. Machine learning models for complex classification problems such as this mainly benefit from larger sample sizes with many predictor variables in order to better train the models. Furthermore, we chose to incorporate sociocultural risk factors for suicide in an attempt to diversify our predictor variables. Unfortunately, it appeared that these variables had little predictive value in our models. In fact, we hypothesize that if more relevant clinical variables were available to us such as psychotic symptom severity, our predictive accuracy would increase. Nevertheless, our models’ strength lies in the robustness of our classifier label in which individual clinical assessments were conducted with both interview and self-report questionnaires along with medical chart reviews to assess patients' history of suicide attempt.

4.6 Conclusion

In the current study we reported the use of two machine learning models with decent accuracy in stratifying a sample of patients with schizophrenia as either suicide attempters or non-attempters using sociocultural, clinical, and demographic predictor variables. Our findings suggest that these machine learning models have some clinical utility in classifying schizophrenia patients according to their risk for suicide. Indeed, machine learning models with binary classification algorithms may one day compliment and improve the diagnostic reliability of many psychiatric phenotypes. By utilizing large scale clinical data and advanced probabilistic algorithms we may be able to develop highly predictive models that will support clinical decisions related to suicide risk. Doing so can improve therapeutic interventions for patients with
schizophrenia by accurately and swiftly identifying patients who may be at the highest risk for a future suicide attempt. Nevertheless, improvements to our model are contingent on a multimodal approach to integrating genetic, neuroimaging, and more predictive clinical variables to accurately predict suicide risk in a cohort of patients with schizophrenia.
Chapter 5

5 General Discussion

5.1 Discussion

Suicide is a major public health concern and one of the leading causes of death in the world. There are many reasons why individuals may decide to attempt suicide, though these are usually contingent upon some form of mental illness (WHO, 2014). Preventing suicide is the ultimate goal of suicide research, which aims to identify the mechanisms behind suicide; whether these are environmental or biological in nature. Unfortunately, suicide can place a heavy emotional and economic burden on those affected (Turecki, 2014), who are termed “survivors”.

The stigma surrounding any form of suicidal ideation and behaviour largely contributes to those at risk not seeking out the treatment that they desperately need. Furthermore, there is an added stigma surrounding mental illness, which is often associated with suicide, and serves to only exacerbate the issue of unsought but much needed help. Particularly in patients with schizophrenia, suicide is a major cause of concern. Those afflicted by schizophrenia often have a lower life expectancy than the average population; a troubling statistic that is largely attributed to increased suicidal behaviour (Saha et al., 2007). Many studies have attempted to uncover and synthesize the available knowledge on the potential risk factors for suicide in schizophrenia. These often result in a putative clinical and psychological profile of an individual with schizophrenia who may be likely to attempt suicide. A patient with schizophrenia who may be considered at high risk for suicide is usually defined as a male in his youth with overall good functioning, who received treatment at a later age, with a relatively high IQ, and falling under the paranoid type of schizophrenia (Fenton, 2000). However, the clinical utility of such measures is often limited and in fact, produces many false positive findings (Pompili, 2007). Therefore, it is
imperative that a means of accurately predicting suicide risk in patients with schizophrenia be developed to prevent potentially avoidable fatalities.

The implications of the work described in this thesis are that we proposed that factors associated with ethnicity and immigration may also contribute to the risk for suicide in schizophrenia. While these factors have been put forward as variables that contribute to both psychosis and suicide independently, no studies have attempted to uncover their role in suicide attempts among patients with schizophrenia. Therefore, the contents of this thesis aimed to support our hypothesis that having immigrated to a new country and being of a visible minority status increases the likelihood of a patient having attempted suicide. Furthermore, as the mechanisms behind suicide attempting are complicated, we realized that perhaps sociocultural factors alone may not be sufficient in reliably predicting suicide attempters. As such, we assessed the efficacy of machine learning models to distinguish between suicide attempters and non-attempters by using sociocultural risk factors, in addition to previously suggested clinical risk factors. We hypothesized that the inclusion of variables pertaining to immigration and ethnicity might enhance the predictability of these models in identifying suicide attempters. The results we described in the previous chapters failed to support our hypothesis that sociocultural variables pertaining to immigration and ethnicity have significant predictive value for suicide attempt history. By testing their association with suicide attempts in a logistic regression model and by including them in more sophisticated machine learning classification algorithms, we failed to find any significant impact of these variables in predicting lifetime suicide attempts. Although our rationale for pursuing these aims appeared sound we were unable to demonstrate that sociocultural variables associated with immigration and ethnicity improve our prediction of suicide in schizophrenia.
5.1.1 Immigration and Ethnicity as Predictors

Previous evidence has suggested that the incidence of schizophrenia may be greater among immigrants and/or individuals of a visible minority status (Kirkbride et al., 2012; Fearon et al., 2006). In particular, the sociocultural factors associated with immigration may be a contributing facet to the increase of psychosis among this population (Li et al., 2012; Cantor-Graae, 2007). Furthermore, studies have shown that the sociocultural factors associated with immigration may increase the risk for suicidal behaviour (Berg et al., 2014; Spallek et al., 2015). Based on the findings that immigration and ethnicity may increase the risk for psychosis and suicide, we attempted to determine whether these factors are predictive of a lifetime history of suicide attempts among schizophrenia patients. We hypothesized that patients who have immigrated prior to the onset of their psychosis are more likely to have attempted suicide during their lifetime. This was tested in multiple parts in this thesis by collecting a sample of schizophrenia patients with clearly defined suicide attempt history and ethnicity according to genetic and self-reported measures. Lastly, we aimed to collect sociocultural and clinical data on individuals with schizophrenia in order to develop a predictive machine learning algorithm for distinguishing between suicide attempters and non-attempters. To our knowledge, this is the first body of research that has attempted to investigate whether factors associated with immigration and ethnicity increase the risk for suicide attempt in schizophrenia. Furthermore, we present the first attempt at using machine learning algorithms to classify individuals with schizophrenia as suicide attempters or non-attempters.

Unfortunately, in our attempt to support our hypothesis using a sample of schizophrenia patients, we were unable to uncover a significant association between suicide and self-reported ethnicity (European Caucasian or Other ethnicity) or immigration status (immigrant versus native-born). Even after using a more refined definition of ethnicity that incorporated
geographical ancestry according to ancestry-informative genetic markers, we similarly found no significant relationship between ethnicity and a history of suicide attempts. On the other hand, the inclusion of these sociocultural factors in our machine learning model supported its ability to distinguish individuals with schizophrenia as suicide attempters or non-attempters with some success. However, the inclusion of these factors had only a marginal effect on the predictive ability of these models, with patient data and clinical scales being the most important features for suicide classification.

The most interesting positive finding from the analysis revealed a significant association between self-reported ethnicity and the number of hospitalizations. We found that among patients with schizophrenia, a higher number of hospitalizations was associated with those who self-reported as either white North American or white European. This finding was only evident with self-reported ethnicity and was not replicated when using genetic ancestry as defined by our STRUCTURE analysis. In an attempt to interpret these results, it is suggested that the lower number of psychiatric admissions among non-European Caucasian patients may be due to cultural and language barriers (Lindert et al., 2008). While we did not find any evidence for such, it is plausible that this link may be exacerbated among immigrants due to increased barriers to healthcare, in addition to being afforded fewer socioeconomic opportunities as a newcomer. Preferential selection by the gatekeepers of healthcare services as well as fear of hospitalization on the part of the immigrant groups may also be factors. On the contrary, a higher number of hospitalizations may be indicative of more severe symptomatology among groups with higher admission rates. In many epidemiological and clinical studies of schizophrenia, a high number of hospitalizations is often an indicator of high symptom severity and is even an important risk factor for suicide in schizophrenia (Togay et al., 2015; Popovic et al., 2014). This would suggest that individuals of a European Caucasian background experience more debilitating symptoms
and/or are more likely to seek psychiatric care. It is important to note, however, that our current sample has a disproportionate representation of European Caucasian individuals. In our analysis of ethnicity and immigration as predictors of suicide attempt in schizophrenia we utilized a sample of 276 individuals, of which 213 (~77% of our sample) were self-reported European Caucasians. This class imbalance may lead to an obvious bias in our findings suggesting that the number of hospitalizations is associated with self-reported ethnicity. Lastly, there has been evidence to suggest that low SES is associated with less frequent use of mental health services (Sripada et al., 2015). In reality, our effect may actually only be capturing this association between SES and the number of hospitalizations. There is considerable evidence to suggest that racial differences in SES are present, whereby European Caucasians may on average be positioned at a higher SES. This racial disparity in SES has been arguably a driving force behind the quality of health care that minorities receive and consequently poorer health outcomes mostly based on less frequent uses of health care services (Williams et al., 2010). Nonetheless, access to health care in Ontario is available to all citizens and permanent residents.

Though this is difficult to clearly ascertain, it is likely that ethnicity plays an important role on the quality and use of mental health services. Ethnicity, however, cannot be considered as a singular factor alone. In correlation with ethnicity, many ethnic groups are defined by many sociocultural factors such as language, religion, and place of birth (Anderson et al., 2015). Because of these unique social groups, nuanced differences exist between groups, which influence their decision to seek psychiatric care (Cauce et al., 2002). One concern that we were unable to control for in our study is the perceived differences in ethnic groups and their interactions with healthcare staff (McGovern et al., 1994). An individual’s ethnic identity may even influence the perceived severity of the psychosis, as many clinicians may be more inclined to consider a poorer diagnosis for immigrants and ethnic minorities due to a lack of sensitivity to
cultural norms (Adeponle et al., 2012). As a result, this practice may likely be a contributing factor for why some studies have reported increased rates of schizophrenia among immigrants. Ethnicity, in itself, is a difficult parameter to define when considering the range of sociocultural variables associated with ethnic identity. However, with the resources available to us, we believe we have captured as much of that identity as possible, as compared to the scientific literature.

5.1.2 Suicide Classification

Over the course of many years of suicide research, the predominant goal has been to elucidate the potential risk factors for suicide attempts. This was primarily carried out by isolating clinically relevant variables that appeared to be more frequent among suicide attempters as compared to suicide non-attempters (Fenton, 2000). While we have come a long way and identified many predisposing factors that clinicians may use to make a judgment as to which patients are at risk for suicide (Popovic et al., 2014; Togay et al., 2015); we continue to rely on sometimes contradictory associations and an extensive review of the individual’s course of treatment. The novelty presented in the contents of this thesis is primarily attributed to our machine learning models for the classification of suicide attempt history in schizophrenia. A number of risk factors have been given for a higher incidence of suicide attempting in schizophrenia. However, we are the first to bring together these various risk factors in order to build a classification model using sophisticated machine learning techniques which are able to use high-dimensional data to make simple classification decisions.

We included predictor variables in our model, mainly according to their availability in our sample, while also considering previously identified risk factors in the literature that are likely to predict a history of suicide attempts. Many important variables suggested in the literature to be associated with suicide in schizophrenia had a greater weight on the predictive capabilities of our
models. Weak associations were found with many sociocultural risk factors which we suggested and included in the model relating to immigration and ethnicity. Essentially, our models were no better when these predictors were included. In fact, with the inclusion of solely sociocultural factors our models’ performance decreases noticeably. Without the use of the CTQ and NEO-FFI clinical scales and important patient demographic data, our models would have relatively poor predictive ability.

The complexity of some classification problems, such as this one, is mainly because of the complex phenotype under investigation. Here, we attempt to classify individuals with schizophrenia as either suicide attempters or non-attempters based on observations of relatively simple predictor variables. Given the diverse range of behaviours involved on the continuum of suicidal behaviour, it is reasonable to assume that there is a great deal of heterogeneity in the suicide attempt phenotype which we failed to capture because of the present state of the classification of suicide attempts. Similar to the protean nature of psychosis, some individuals may display suicide attempts in many different ways. By broadly categorizing individuals as suicide attempters solely according to whether they have committed a self-injurious behaviour with some intent to die, we miss the diagnostic nuances that may further stratify suicide attempters. Prevalent within schizophrenia, many suicide attempts occur without clear warning signs or even rational intentions. Because of the nature of schizophrenia, many patients may experience bizarre delusions or sensory hallucinations (Togay et al., 2015) which lead them to attempt suicide but for clearly different reasons than perhaps for other psychiatric patients or the general population.

Suicide risk prediction has been a goal of suicide research for many years. As early as in 1983, well before the advent of sophisticated machine learning algorithms, researchers were attempting to use statistical methods to predict psychiatric patients who were likely to attempt
suicide (Pokorny, 1983). This study by Pokorny (1983) used a large clinical sample of 4,800 psychiatric in-patients to assess whether they would be likely to attempt suicide in the future. Using a wide range of clinical items and instruments, the author conducted an item-by-item analysis to determine their prediction scores in determining which patients were at risk for completing suicide in the future. Ultimately, this study did poorly in accurately identifying individuals who went on to complete suicide, as this was a prospective study that followed-up on patients after 5 years. Using this discriminant function analysis, only few patients were accurately identified and many false positive predictions were made. Now that we have access to more sophisticated algorithms and computational capabilities, many researchers are attempting to continue this idea of predicting suicide in vulnerable populations. Recently, as the only other study to have focused specifically on suicide attempts, Passos et al. (2015) attempted to identify a clinical signature for suicide attempts in a sample of individuals diagnosed with a mood disorder. The model that was able to classify suicide attempters and non- attempters with the most degree of success using the RVM algorithm. Our classifiers performed quite similarly, with Passos et al. (2015) having slightly more success in distinguishing attempters from non- attempters. While it is much too early to say, and many more studies of this kind must be done, it appears at first glance that perhaps correctly classifying suicide attempters is a difficult task with the current predictor variables we commonly consider. Despite having a larger sample size, different predictor variables and alternative classification algorithms, our models performed acceptably. This suggests that perhaps the suicide attempt phenotype is quite complex and difficult to predict due to high variability in factors that lead to this behaviour. We suggest that novel and more predictive risk factors must be uncovered and included in models of this kind. Nevertheless, we performed the first study to use classifiers in order to distinguish suicide attempters from non-attempters among a clinical sample of schizophrenia patients.
5.2 Limitations

The study design of our experiments was chosen considering available data collected during cross-sectional clinical assessments with the goal of testing the main hypothesis. Nevertheless, the results of the experimental procedures must be interpreted in the context of several limitations. As our hypothesis considered ethnicity and immigration as important risk factors for suicide attempt in schizophrenia, it was important that we carefully define their terms. Initially, we attempted to test whether ethnicity defined according to self-report may predict differences in lifetime suicide attempt history in a sample of schizophrenia patients. When considering ethnicity according to self-report, this may incorrectly be defined upon cultural or even delusional ideation. A portion of the patients in our sample, due to psychotic beliefs, reported belonging to an ethnic group that they were in reality not a part of. This was confirmed by comparing their responses between our self-reported ethnicity questionnaire (see Appendix A) and our grandparent ancestry questionnaire (see Appendix B). Furthermore, we wanted to ensure that ethnicity was defined according to geographical location in addition to sociocultural identity, to control for any potential genetic differences. By considering genetic ancestry and self-reported ethnicity we found 29 individuals who self-reported as being non-European Caucasian but clustered in the European ancestry sample. In reality, the number of ancestry-informative genetic markers we had available to categorize individuals according to ethnicity were not enough to sufficiently stratify individuals into discrete ethnic clusters. In traditional genetic association studies, having individuals of a non-homogenous ancestry will lead to spurious findings. Therefore, particularly in large GWAS studies, thousands of ancestry-informative markers are commonly used to accurately stratify individuals according to ancestry. The difference in using a larger number of ancestry-informative markers is evident even when comparing Figure 2-1 and Figure 3-1. With a larger number of SNPs, the MDS approach is better
capable of clustering individuals according to the three HapMap reference populations. Unfortunately, our multiple choice self-report questionnaire is more prone to discordant ethnicity according to identity and genetic ancestry, and also presents issues in allowing patients to correctly select the appropriate ethnicity. Our self-report questionnaire provides a number of options for participants to select their identified ethnicity; leading to more confusion as to which category to select. Because of the variety and number of options to select from, many patients experienced difficulties in selecting what they believed was the appropriate ethnicity. On the other hand, we classified genetic ancestry according to only three major ethnic groups (European, Asian, and African) which does not adequately capture the diversity of human populations. In addition, as confirmed in our attempt to validate our self-report and grandparent questionnaire to MDS-defined ancestry, we found that the self-report questionnaire was prone to more errors. By giving the participant the opportunity to select an ethnicity from a given list as opposed to asking them about their grandparents geographical ancestry, we allowed them to select an ethnicity that may not be accurate according to their true identity or of which may have been chosen due to some quasi-delusional ideation.

A major limitation of our study is that it is difficult to isolate the effect that either ethnicity or immigration had on suicide attempt risk. Because immigrants often also belong to a visible ethnic minority group, it is difficult to claim that one factor had more of an effect than the other. Both variables often work in tandem with one another to produce the stressful circumstances that often lead to psychosis and suicide. As a result, we tested the effect each variable had independently had on predicting a history of suicide attempt and also their interaction in our logistic regression model. In addition, by grouping all immigrants in our sample as a single group, we inevitably lost nuanced differences in the stressors and life experiences that these individuals faced. Unfortunately, our sample of 276 participants only
contained a total of 64 immigrants, which accounts for only 23% of the sample. Therefore, having segregated these individuals even further would lead to a severe reduction in our ability to detect any significant associations between immigration and suicide attempts in schizophrenia. Indeed, the course of migration that the participants’ experienced is an important clinical factor which should be accounted for. Another major limitation of grouping immigrants as a single entity is that we were unable to assess the effect of having refugee or asylum seeker status on suicidal behaviour. Because immigrants who fall under these categories often have markedly different (and often worse) experiences when leaving their native country and adapting to their new host country, they may have been at a greater risk for suicide. Lastly, because we considered participants as immigrants only if they had immigrated to Canada during their lifetime, we essentially only considered first-generation immigrants. Unfortunately, because the nature of this study was cross-sectional and retrospective in its study design, much of the sample was collected before the beginning of this project. At the time of data collection, we had not considered including second-generation immigrants as well. For the sake of maintaining a suitable sample size we opted to focus our analysis solely on first-generation immigrants with strong clinical data. Overall, had we included these important caveats involved in immigration in our analysis we might have been better suited to identify a relationship.

In our analysis of machine learning models for the classification of suicide attempters and non-attempters there were several limitations that influenced our results. The major limitation in our analysis is our choice of predictor variables. As our main hypothesis was to assess the effect of immigration and ethnicity in predicting a history of suicide attempts in our sample of schizophrenia patients, we elected to include these variables in our models. Unfortunately, it appeared that these variables did not have much of an effect on predicting suicide attempters in our sample. Therefore, a major limitation of our models was that we largely relied on only these
sociocultural risk factors and pertinent clinical information. Had we had access to a larger number of clinical variables and diagnostic items, we believe that our prediction accuracy would have benefited greatly. Furthermore, if we had access to multimodal data such as imaging or genetics, we believe that our prediction accuracy and model performance would have greatly improved. The benefit of using a regularization method such as LASSO is that it enables users to select a large number of variables while not leading to overfitting. By not having important risk factors that were repeatedly found to be associated with suicide attempting in schizophrenia in the literature, we cannot confidently say that our model is a true predictor of suicide attempters. Many variables such as marital status, symptom severity, treatment at the time of the suicide attempt, and education were not included in our model and thus limited our ability to make meaningful classifications. In addition, our relatively small sample size inhibited us from making true predictions from our data. Unfortunately, the numbers of suicide attempters and non-attempters in our sample were skewed in that there were more non-attempters present. By not having as many suicide attempters, our algorithms were less able to 'learn' from the training data to make accurate predictions in the 'test' data. Therefore, a major limitation of our analysis was the exclusion of many crucial predictor variables that are known to be associated with suicide in schizophrenia.

Lastly, there are many key limitations that were unavoidable. Many of the participants included in our study were recruited as out-patients in the community. While some participants were in-patients at the time of their assessment, the large majority were not. Therefore our analyses are inherently biased towards patients with better functioning and potentially less severe symptomatology. As a result, the findings of our analyses may not be applicable to the general population with schizophrenia. In addition, there is a great deal of heterogeneity in our sample as many patients experience a wide variety of symptoms and many of them may have been
comorbidly diagnosed with a mood disorder or schizoaffective disorder. Depressive and/or manic episodes, which often characterize schizoaffective disorder, may also be important contributors to an increase risk for suicide. Therefore, by aggregating schizophrenia and schizoaffective patients together we may be masking specific risks for these two syndromes independently. Furthermore, our sample is not representative of the larger population, in that many participants were recruited solely in Toronto at the Centre for Addiction and Mental Health and its satellite clinics. As we’ve seen in many studies on migration and suicide, the host population and country play a vital role. It is imperative to note that our findings may only be applicable to the Toronto area and perhaps Canada as a whole, although many differences in ethnic densities and societal norms exist within the country. Finally, our cross-sectional study design is a major limitation in that the data we collect is primarily retrospective and relies on patient testimonies of their experiences. Without having assessed these patients at the time of their suicide attempt or following their migration to Canada, we may be missing valuable information that may be lost due to recall bias. To obtain the most accurate and useful information regarding these suicide attempts, it would be ideal to assess these patients near the time of their suicide attempt. Of course, this is a difficult endeavor and not a reasonable aim given the small sample size we would obtain.

Despite these limitations, this work is the first of its kind to attempt to analyze whether suicide attempts in schizophrenia are a result of sociocultural factors associated with immigration and ethnicity. Furthermore, we provide the first study to utilize a classification algorithm using machine learning to identify the likelihood that patients in our sample are suicide attempters or non-attempters.
5.3 Conclusions

The body of work described in this thesis illustrates our approach to testing whether the risk for suicide attempting among schizophrenia patients is mediated by sociocultural factors involved in immigration and ethnicity. We tested our hypothesis by recruiting a well-characterized sample of schizophrenia patients and analyzed the likelihood that being an immigrant or belonging to a visible ethnic minority increased the risk for lifetime suicide attempts. Through our analyses we were unable to support our hypothesis, reporting non-significant results for this association. Furthermore, we tested the contribution of these sociocultural risk factors in a machine learning classifier to predict whether patients with schizophrenia are suicide attempters or non-attempters. Unfortunately, the aforementioned limitations thoroughly described make it difficult to accept our hypothesis altogether.

In our first investigation (Chapter 2), where we attempted to validate self-report measures of ethnicity in a psychiatric population, we found that asking for the geographical location from where the participants’ four grandparents originated was the most consistent with ethnicity defined by MDS, which is considered a gold standard for testing population stratification. Many GWAS employ the MDS when searching their genetic sample for individuals who are not of a clearly defined homogenous ethnicity. This is essentially to prevent significant results that are only due to population stratification. Here we argue that by using the information regarding the geographical ancestry of the participants’ four grandparents, researchers may avoid recruiting participants who do not fit their criteria of a homogenous population. By administering this questionnaire asking for grandparents’ ancestry, researchers can avoid the unnecessary loss of financial and human resources in recruiting, assessing, genotyping and testing these individuals who will ultimately be discarded at the MDS stratification stage. The goal of this research was to support large-scale genomic studies in order to maximize efficiency and improve the validity of
findings from these association studies. Particularly in the psychiatric population, patients may not always be reliable in correctly identifying their true ethnicity due to the positive and cognitive symptoms associated with schizophrenia. In summary, collecting self-reported ethnicity from psychiatric patients may not always provide reliable data. Therefore, we proposed that identifying the geographical ancestry of participants’ four grandparents offers a more efficient alternative to prevent lost resources on individuals who will ultimately be excluded.

In our second investigation (Chapter 3), we attempted to demonstrate whether immigration and ethnicity were predictive of a history of suicide attempts in schizophrenia. Unfortunately, we found that being of a visible minority group and having immigrated during one’s lifetime might not be predictive of previous suicide attempts. Interestingly, we were able to show that self-reported ethnicity does in fact have an association with the number of hospitalizations in our sample of schizophrenia patients. We suggest that these findings are perhaps related to the accessibility of mental health services. It is suggested that European Caucasians are more likely to request or be compliant with psychiatric hospitalization as compared to ethnic minorities due to cultural perceptions of mental health. In this study, we also attempted to demonstrate a robust definition of ethnicity by combining self-reported ethnicity with genetic ancestry for our analysis. We utilized the MDS with ancestry-informative genetic markers to obtain the geographical ancestry of our participants. This was compared to what participants’ selected through self-report. We then conducted our analysis using only individuals whose self-reported ethnicity matched their geographical ancestry. Doing so allowed us to reasonably remove doubt as to whether cultural ethnicity or biological ancestry was affecting our findings. In the end, however, we were unable to demonstrate a relationship between ethnicity and suicide attempt history in schizophrenia.
In our last investigation (Chapter 4), we attempted to use two machine learning algorithms to test their accuracy in classifying schizophrenia patients as either suicide attempters or non-attempters. Our models were based on sociocultural factors associated with immigration and ethnicity, and clinical predictor variables. Looking at our results, we were able to demonstrate that there is some clinical utility in our models for predicting which individuals had a history of suicide attempts. Given the predictor variables we used, we were able to show decent accuracy and both models showed almost equivalent AUC values. The binary classification of machine learning models one day assist clinicians in identifying patients who are at the greatest risk for suicide using only the clinical variables collected during psychiatric assessments and evaluations. Improvements are still necessary in selecting the appropriate risk factors and predictor variables for these algorithms, and accurately identifying the suicide phenotype. Nonetheless, there is great promise for the use of these algorithms in psychiatric research and eventually in therapeutic practice.

The studies presented in the contents of this thesis have established the first investigations into the relative contribution of sociocultural risk factors, related to migration and ethnicity, in predicting the risk for suicide attempt in schizophrenia. While we were unable to provide significant results for our hypothesis, it is our hope that future studies attempt to replicate the findings presented here and improve upon the models that we have suggested.

5.4 Future Directions

5.4.1 Longitudinal Assessment of Immigration

As this project demonstrates the first attempt at identifying whether immigration influences the risk for suicide attempt in schizophrenia, there are notably areas of improvement that future studies can implement. As discussed in the limitations, our classification of
immigrants was grossly oversimplified. Future studies aiming to replicate our results should address the need to have more nuanced descriptions of the immigrants in the sample according to their status. As we only collected information regarding first-generation immigrants, future work must aim to consider the effect of being a second-generation immigrant on the risk for suicide in schizophrenia. Furthermore, it is advisable that immigrants who arrive as refugees or asylum seekers also be noted as a covariate in the analysis. Individuals coming from war-torn countries or unsafe zones are more likely to experience post-traumatic stress and difficulties adjusting to a new life and environment which they most likely did not choose willingly. Immigrants to Canada arrive from many different countries around the world. The admission policies in place favour immigrants with professional skills who are able to positively contribute to society. There is also a large number of refugees that comprise the immigrant population (Anderson et al., 2015), however a limited number of studies have actually investigated the unique risk for psychosis and suicide that refugees face. In contrast to traditional immigrants, refugees differ in their level of education, prior SES, and their overall pre- and post-migration experiences due to the trauma experienced in their country of origin and their reason for leaving (Anderson et al., 2015). As mentioned previously, a history of childhood trauma has a significant influence on the risk for suicide (Hassan et al., 2016). Refugee groups experience greater adversity than the average immigrant in terms of securing housing, accessing health care, and economic opportunities (Beiser, 2005). The post-immigration phase is a sensitive period for immigrants’ mental health in which refugees experience comparatively worse hardships that further add to the pre-migration stress that led them to flee their country of origin in the first place (Porter & Haslam, 2005).

A major improvement in collecting adequate data for the purpose of this study would be to have a much larger recruitment effort. As our sample is recruited entirely from a single institution and city, future work would benefit from having a multi-site recruitment strategy,
which aims to collect participant information from patients throughout the province or country. Not only does this have the benefit of significantly increasing sample size and the number of immigrants in the sample, it ensures that any results obtained are not specific to the area in which the participants were recruited. Having a more diverse sample allows researchers to be more confident in claiming any generalizability of their findings to other patients with schizophrenia.

As a further improvement to recruitment strategies, it would be advisable that the study be redesigned such that individuals are assessed closer to the time of their first suicide attempt. As many large research hospitals have individuals dedicated solely to the recruitment of a specific population (i.e. research analysts), it would be a major improvement if future studies could assess patients almost immediately following a suicide attempt. Of course, this is a challenging feat, but if successful, the data would be much more rich, reliable and predictive. By assessing the patient near the time of the suicide attempt, researchers would be able to gather key information such as the medication and dosage of the patient at the time of the attempt; the circumstances surrounding the decision to attempt suicide; a more accurate age at which the attempt occurred; and an accurate description of the details involved in the actual suicide attempt (including lethality and intent).

5.4.2 Comparison of Clinical Judgment and Machine learning

Machine learning algorithms offer the unique approach of being able to synthesize a vast array of variables in helping to predict different classification labels. While our models failed to incorporate more than just sociocultural and clinical variables, this is still very much possible. Future studies of this kind, that aim to replicate our work and improve on the models we provide, would be well suited to increasing the number of relevant predictor variables in the model. Indeed, many of the sociocultural predictors we used had very little effect, if any, on identifying
suicide attempters or non-attempters in our sample. By incorporating multidimensional data, these algorithms may be able to 'learn' from a variety of predictors so as to best decide on the classification of schizophrenia patients’ suicide attempt risk. Ideally, if the data becomes available to researchers with a similar goal to ours, genetic or imaging data should be included in their models as well. Furthermore, there should be an increased effort to include many, if not all, of the risk variables described in the literature over time. By having the most predictive variables in the model, these machine learning algorithms will undoubtedly improve their prediction accuracy. If so, these algorithms could one day serve a clinical purpose in assisting clinicians in being able to identify those with schizophrenia at the greatest risk for future suicide attempts.

In addition, I suggest that these algorithms be compared to actual clinician judgments in identifying individuals who are likely to be future suicide attempters or non-attempters. Assessment tools to evaluate the risk for suicide mainly recognize demographic, diagnostic, and health-related risk factors. However, it has been mentioned in systematic reviews that there is a lack of prospective studies that aim to evaluate the effectiveness of these risk assessment tools (O'Connor et al., 2013). This is a major concern in retrospective studies such as ours. While we may evaluate the efficacy of our model internally, it is difficult to say how these machine learning models would hold up when used on an external sample of schizophrenia patients and how predictive it would be to identify individuals who are likely to attempt suicide before an event even occurs. Therefore, I suggest that our machine learning models be applied to samples outside of our own in an attempt to replicate their accuracy and predictive value. If these models provide similar summary statistics as to the model's efficacy, it is plausible to assume that these risk factors and models are rather robust in classifying suicide attempters from non-attempters. Furthermore, it would be interesting to compare the predictive accuracy of these models to clinician-based judgment on the risk for suicide. I propose a study design which involves training
a classifier on clinical and demographic risk factors found in patient records that pertain to suicide attempters. On a new set of medical records with the same categories of clinical; demographic; diagnostic information; the classifier would be tested to assess how accurately it may be able to discern the records of a suicide attempter or non-attempter with schizophrenia. Similarly, a group of clinicians would also have access to the same medical records and be asked to predict who they would discern as likely to be suicide attempters or non-attempters. This retrospective study would then compare the success rate of both the machine learning classifier and the clinician's judgment. By performing this comparison, we could assess the utility of these algorithms in a clinical setting as a valid risk assessment tool for suicide in schizophrenia and potentially other psychiatric disorders in the future.

5.4.3 Improved Phenotype for Classification

One of the prevailing difficulties in suicide research is the need to clearly define the suicidal behaviour under investigation. As the purpose of this thesis is to solely assess the risk for suicide attempt as its own distinct phenotype, we must fully understand the variability that this behaviour entails. There is a great deal of variability among suicide attempters, mainly among the methods that they choose to end their lives with; the lethality of such measures; the reasons for attempting (psychotic or depressive); and also the number of attempts (one attempt vs. multiple). Because suicide attempting on its own ranges on a spectrum, it makes it difficult for algorithms to correctly identify suicide attempters as they are difficult to neatly compartmentalize as their own distinct class. For example, it does not seem reasonable to assume that someone who attempted suicide once had the same level of risk as someone who attempted more than ten times. The simpler the classification label, the easier it is for the machine learning classifier to correctly distinguish between two groups. Therefore, we suggest that future studies
aim to refine their definition of suicide attempt using more than just the general yes/no response. In theory, someone who attempted suicide once under slightly lethal means with limited intent may be no different than someone who has never attempted suicide though came very close to it due to severe suicidal ideation.

Having the classification labels be stratified to more extreme ends only further helps the classifier distinguish between the two groups. As an improved method of defining suicide attempters in future algorithms of this kind, we propose that future studies classify individuals with schizophrenia as either non-attempters or "serious" suicide attempters. Suicide attempters, in this case, would ideally be those who are more extreme in their lethality, suicidal ideation, and intent. These would be characterized by collecting information regarding the suicide methods used in their attempts and classifying them as either violent (i.e. hanging, use of firearms, cutting) or non-violent (i.e. overdose); and assessing the lethality of their actions using the lethality item in the Columbia Suicide Severity Rating Scale (C-SSRS). Furthermore, the intent of the suicide attempt can be assessed using the Beck Suicide Intent Scale (SIS). Lastly, suicidal ideation can also be assessed using the C-SSRS, as well. By calculating the mean values of these score across a sample of suicide attempters with schizophrenia, only those that score above a specified threshold for lethality, intent, or ideation (depending on what the author deems the most severe) may be included in the classifier as suicide attempters.
References


Hassan, A.N., Stuart, E.A., & De Luca, V. (2016). Childhood maltreatment increases the risk of suicide attempt in schizophrenia. *Schizophr Res* [Epub ahead print]


Appendices

Appendix A: Self-Report Ethnicity Questionnaire

Study ID: ___________ Visit #: ___________ Visit Date (DD/MM/YYYY): __________

Genetics of Affective Disorders
Self-Report Information Sheet

Please indicate your ethnicity below by circling the most appropriate response. Thank you!

ABO  Aboriginal  ICA  Indian Caribbean
AEA  Asian East  MBA  Mixed Background
ASO  Asian South  MEA  Middle Eastern
ASE  Asian South East  OTH  Other
BAC  Black Caribbean  UNK  Unknown
BAF  Black African  WEA  White European
BNA  Black North American  WNA  White North American

Additional details (if needed): ____________________________________________________________________________
_________________________________________________________________________________________________________
# Appendix B: Grandparent Ethnicity Questionnaire

Ethnic Status: (Where did the subject’s family come from?)

<table>
<thead>
<tr>
<th></th>
<th>Maternal Grandmother</th>
<th>Maternal Grandfather</th>
<th>Paternal Grandmother</th>
<th>Paternal Grandfather</th>
</tr>
</thead>
<tbody>
<tr>
<td>European/Caucasian</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>African Descent/</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>African American</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Indian</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Caucasian (e.g.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistani, Indian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>(e.g. Chinese,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Native N. American</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Other:______________</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Your Religion: ________________

Your Place of Birth:

Mother’s: ____________________

Father’s: ____________________

Primary language: ____________