University of Ottawa

Risk factors for suicidal behaviour among Canadian civilians and military personnel:

A recursive partitioning approach

by

Corneliu Rusu

A thesis submitted to the Faculty of Medicine in partial fulfillment of the requirements for the degree of

Master of Science

in

Epidemiology

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

Abstract

Background

Suicidal behaviour is a major public health problem that has not abated over the past decade. Adopting machine learning algorithms that allow for combining risk factors that may increase the predictive accuracy of models of suicide behaviour is one promising avenue toward effective prevention and treatment.

Methods

We used Canadian Community Health Survey – Mental Health and Canadian Forces Mental Health Survey to build conditional inference random forests models of suicidal behaviour in Canadian general population and Canadian Armed Forces. We generated risk algorithms for suicidal behaviour in each sample. We performed within- and between-sample validation and reported the corresponding performance metrics.

Results

Only a handful of variables were important in predicting suicidal behaviour in Canadian general population and Canadian Armed Forces. Each model's performance on within-sample validation was satisfactory, with moderate to high sensitivity and high specificity, while the performance on between-sample validation was conditional on the size and heterogeneity of the training sample.

Conclusion

Using conditional inference random forest methodology on large nationally representative mental health surveys has the potential of generating models of suicidal behaviour that not only reflect its complex nature, but indicate that the true positive cases are likely to be captured by this approach.

To my family

Remember that all models are wrong;

the practical question is how wrong do they have to be

to not be useful.

George Box

Acknowledgments

I owe my deepest gratitude to Dr. Ian Colman who opened me the door to the realm of machine learning and made this thesis possible. To my colleagues Dr. Bryan Garber and Dr. Mark Zamorski, thank you for your constructive advice and encouragements through the makings of this project. I also would like to thank Dr. Ronald Kessler for his patience in answering my pestering questions regarding his team's pioneering work of applying machine learning in suicide research. Likewise, I would like to thank Dr. Max Kuhn, Dr. Robin Genuer, and Dr. Torsten Horton for their insightful inputs related to the use of the R packages they authored. Lastly and most importantly, I would like to thank my family for their unwavering love and support.

Table of Contents

List of abbreviations

AUC	Area under the curve
CAF	Canadian Armed Forces
CCHS-MH	Canadian Community Health Survey – Mental Health
CFMHS	Canadian Forces Mental Health Survey
GAD	Generalized anxiety disorder
MDE	Major depressive episode
OOB	Out-of-bag
PTSD	Post-traumatic stress syndrome
VIM	Variable importance measure

List of tables

List of figures

1. Introduction

1.1 Epidemiology of suicide

Suicide is a major public health concern in Canada and internationally, with devastating and far-reaching impact on family, friends, and communities. Every 2 seconds someone attempts suicide worldwide [1] and every 40 seconds a death by suicide occurs [2]. Suicide affects people of all ages and backgrounds accounting for half of all violent deaths in men, and nearly three quarters of all violent deaths in women [2].

At the national level, suicide deaths claim the lives of close to 4,000 Canadians each year [3-5]. Over 90% of them were living with a mental disorder at the time of their death [6-7]. For every death by suicide, many are profoundly affected by the loss, struggling with grief in search for solutions. Past-year non-fatal suicide attempts, which outnumbered completed suicides by a factor of 30 in 2013 [8], add to the public health burden of suicide, by their socio-economic impact on communities, due to the utilization of a complex array of health services necessary to treat the survivor. Furthermore, for every attempter there were approximately 5 others who reported having serious thoughts of suicide during the past year [8,9].

Suicide is also the second leading cause of death of currently serving and released Canadian Armed Forces (CAF) personnel, accounting for a quarter of all deaths [10]. Unsurprisingly, a substantial amount of public interest has been paid to this issue, especially if the suicide cases were thought to be related to deployment. A recent cluster of suicides among previously deployed CAF members and veterans has been just one of such instances in which the public, Members of Parliament and military leaders have questioned whether the CAF is seeing increased military suicide [11].

Historically, suicide rates among military personnel have been substantially lower than those of similar age and sex in the general population [12,13]. However, suicide rates have increased dramatically among the US military personnel since the initiation of military operations in Afghanistan and Iraq [14,15], to the point of exceeding those observed in the U.S. general population for the first time in history [16]. A similar pattern has been observed in the CAF: the overall suicide rate slightly increased since 2010, but without exceeding the rate observed in the Canadian general population [13,17,18]. On the other hand, past-year non-fatal suicide attempts were 5 times more prevalent in CAF Regular Force personnel than in full-time employed Canadian civilian population in 2013 [8].

1.2 Risk factors for suicidal behaviour

A broad range of risk factors has been found to be associated with suicide and suicidal behaviour in civilian and military samples. Among them, mental disorders, previous suicide attempts, severely stressful life experiences including traumatic life events, low levels of social interaction and support, and socio-demographic factors have shown the most robust association with suicide behaviour outcomes in previous clinical and population-based studies [19-22].

1.2.1 Psychopathology

As demonstrated by psychological autopsy studies, a history of mental disorder can be traced to most individuals who die by suicide, with over 90% of suicide victims or attempters having had a diagnosable mental disorder around the time of their death [1,6,23,24]. Even the remaining

10% with apparent no detectable mental disorder had subclinical psychiatric symptoms, implying that mental disorder or symptoms are necessary but not sufficient causes of suicide [7]. Mood disorders are associated with about 60% of suicides [1]. Study upon study has demonstrated that individuals afflicted by mental disorders are at an increased risk for suicide and suicidal behavior in both civilian [25-39] and military [15,38,40] samples. Evidence from civilian studies suggests that while mood is associated with suicide ideation, mental disorders defined by agitation, such as generalized anxiety disorder (GAD), social phobia, posttraumatic stress disorder (PTSD), bipolar disorder, substance use disorders or psychotic disorders are more likely to be associated with suicide attempts [36,37]. However, data from the CAF Medical Professional Technical Suicide Review reports shows that more than 40% of suicide victims in 2015 had a documented depressive or substance use disorder, close to 30% had anxiety disorder, while more than 20% had PTSD [18]. 2/3 of individuals who died by suicide in 2015 had comorbid mental disorders diagnosed before their death [18]. Indeed, the risk of suicide is further increased if psychiatric comorbidity is present in the same individual [25,26,41,42]. However, the great majority of psychiatric patients do not die by suicide [20,21,35-37].

1.2.2 Previous suicide attempts

Previous suicide attempts are regarded as one of the strongest predictors of future suicide attempts [26,43-48]. Individuals with such a history are 40 times more likely to die by suicide than those without any attempt [27], and their increased risk for future attempts is likely to persist over the entire adult lifetime [49]. Similar findings have been found in military samples as well: military personnel with a history of pre-recruitment suicide attempts were 4 to 6 times more likely to commit an attempt after joining the military [50,51].

1.2.3 Negative life events

Severely stressful life experiences, such as unemployment [52-54], financial difficulties [52,53], loss through death or separation [52-59], interpersonal conflict [60,61], chronic pain [62-64], physical illness [52-55,65], childhood abuse [66-76], military sexual trauma [77,78], and combat exposure [69,79] are associated with an increased risk for suicidal behavior. Among early negative life stressors, a history of adverse childhood experiences has been shown to have one of the strongest associations with suicide ideation, plans, and attempts [39,70-76]. Virtually all types of childhood abuse (emotional, physical, sexual, exposure to intimate partner violence) increase the risk for suicide in both civilian [75,80-82] and military populations [76]. Moreover, the risk remains significant after controlling for co-occurring psychopathology [75,80]. As for more proximal stressors, recent interpersonal negative life events involving romantic relationships have been found to increase the risk for suicide behavior. There is evidence that these events act as triggers for suicide attempts in the immediate aftermath of their happening [61]. Generally, military personnel are prone to the same family-related and personal stressful life events as their civilian counterparts; however, they face many additional military-related stressors, such as prolonged deployment, combat exposure, repeated deployments with re-exposure, military sexual trauma, physical injury, and traumatic brain injury [20]. A dose-response relationship between the number of traumatic life events and the risk for suicide has been observed in both civilians [75,76,83] and military [69,76].

1.2.4 Social connectedness and social support

Social connectedness and social support, through their various negative facets, such as lack of perceived social support [60], self-reported loneliness [53,84,85], living alone and having low levels of social interaction [60,86], loss of a spouse through death or divorce [52-59], interpersonal conflict [60,61] are considered among the most reliable predictors of suicidal behavior across the lifespan. On the other hand, religious affiliation [86,87], marriage [55,56,86,88], number of children [56,89] and friends are protective against suicide behavior. The same protective and deleterious factors are potentially at play in military personnel [90].

1.2.5 Socio-demographic factors

Socio-demographic factors associated with suicide in the general population (i.e. male sex, younger age, lower education, lower income, unemployment) are largely similar to those found in military personnel [38,91]. However, differences exist. Notably, Canadian women are more likely to attempt suicide than men [92], both in the general population and in the military [38], while the suicide rates in men are nearly four times higher than the rate among women [92]. The difference seems to come from the fact that men often use more lethal means [93], such as firearms or hanging, to end their lives, while women are more likely to choose methods with a greater chance of intervention that will eventually save their lives (i.e. overdose of pills) [92]. Moreover, suicide attempts in the military are more likely to result in death than in the general population, due to their extensive training, familiarity and access to firearms [93].

1.3 Theoretical models of suicide

Surprisingly for a phenomenon responsible worldwide for almost 1 million lives each year [2], and despite a large body of research, the theoretical developments regarding suicidal behavior have progressed somewhat slowly when compared with other areas of psychopathology or health risk behaviour research. Various theories have been proposed in the past 120 years, pertaining to the biological [29], sociological [94], psychodynamic [95,96], cognitive-behavioral [97-99] or interpersonal-psychological [21,100] aspects of suicidal behaviour. Focusing on isolated risk factors, or on relationships between small sets of risk factors, these theories cannot fully explain all the aspects of suicidal behaviour [101]. Many of them are inconsistent with the empirical evidence regarding this phenomenon because they assume that the risk for suicide is elevated only through increasingly severe levels of desire for suicide [21]. Lacking distinct mechanisms through which both suicide ideation and attempts are accounted for, most theories are at odds of explaining the gender distribution or the prevalence of suicidal behavior [21,100]. In other words, they are unable to explain why most individuals experiencing a certain risk factor never engage in suicidal behaviour, much less die by suicide [21, 101], while those who eventually die by suicide have a constellation rather than one or few individual suicide-related risk factors. Furthermore, treating various aspects of suicidal behavior as a single unitary construct, these models of suicide have difficulties explaining that, while there is an important overlap of the risk factors related to ideation and attempts, there are essential differences among them that cannot be ignored [20,21]. Most critical perhaps is the fact that these theories emphasize the role of a particular construct or set of constructs without specifying which facet of suicidal behaviour they are supposed to explain [102]. In the words of Klonsky and May, these "theories of suicide have conflated the question of why people *feel* suicidal with the question of why people *act* on suicidal thoughts" [102]. An overview of the main pros and cons of each current theory of suicide is given below.

1.3.1 Strengths and weaknesses of current theories of suicide

Durkheim's sociological theory asserts that a disturbed regulation by the society in which one lives is the main driving force leading to suicide. In other words, societal forces (social integration and moral regulation), rather than individual factors, are the primary cause for suicide. As the first comprehensive theory of suicide, Durkheim's theory can provide potential explanations for the seasonal variations of suicide, and make associations between specific societal phenomena and shifts in rates over time, at the cost of minimizing individual level factors responsible for suicide [21,103,104]. In other words, it cannot explain why the majority of people does not die by suicide when societal forces are propitious for suicide to occur.

Beck's hopelessness theory asserts that overwhelming, immutable hopelessness is the main cause of suicidal behavior. Several studies showed that, indeed, hopelessness may be a better predictor of suicide than depression alone [97,98]. However, many people who feel hopeless about the future do not attempt suicide. Indeed, by its own, hopelessness has poor specificity [105], being unable to account for the all variance in suicidal ideation and behavior [103,104].

Baumeister's escape from self theory proposes that suicide is an attempt to escape from aversive self-awareness and intense negative emotions [95]. It consists of 5 steps: (1) falling short of standards due to unrealistically high expectations and/or negative life events, (2) internal attributions of self-blame and low self-esteem, (3) heightened sense of self-awareness followed by, (4) negative affect and, (5) cognitive deconstruction to escape negative affect, which results in (6) four consequences of deconstruction that are relevant to suicide: disinhibition, passivity, lack of emotion, and irrational thoughts [103,104]. By involving a lengthy chain of events, the theory correctly implies that suicide is a rare event. However, this is also its main drawback, explaining the paucity of research on the theory: no studies linking causally the 6 stages are available yet [103,104].

Linehan's emotion dysregulation theory posits that suicide is a result of the interaction between an emotionally invalidation environment and emotion dysregulation (inability to properly adjust own emotions), which generates intense negative affects that can be overcome only through extreme behaviours, such as suicide [99]. Empirical support for Linehan's theory is mainly indirect, through studies evaluating the efficacy of dialectical behaviour therapy, a treatment for borderline personality disorder that was derived from the theory and targets emotion dysregulation [103,104]. However, Linehan's model suffers from the same low specificity as other theories, because emotion dysregulation is a common underlying construct in many mental disorders, and most people experiencing them do not engage in suicidal thoughts or behaviours [103,104].

Shneidman's psychache theory asserts that suicide is a consequence of *psychache*, an intense and deep psychological pain that overtakes the mind [96]. Psychache emerges from a profound deficit in one's psychological needs, which are thwarted or distorted, and it will increase the risk for suicide if it becomes unbearable. Moreover, as the individual runs out of options to decrease psychache, the desire for suicide increases, resulting in more severe attempts, with lower chances for survival [103]. Shneidman also lists 7 basic psychological needs for which an individual must have a deficit in order to experience psychache: affiliation, counteraction, defense, inviolacy, shame-avoidance, succorance, and need for order and understanding. Empirical evidence suggests that psychache indeed predicts suicide ideation, however, no studies support the assertion that all the 7 deficits must be met to engage in suicidal behaviour [103].

Joiner's interpersonal-psychological theory of suicide asserts that in order to die by suicide, an individual must manifest both the desire to die and the capability to carry out a suicidal act [21,100]. It is the first theory that uses an ideation-to-action framework, in which different explanations are given for suicide desire and the capability to act on this desire [102]. We refer to it in greater detail in the next section.

O'Connor's integrated motivational-volitional model of suicidal behaviour is a broader theory of suicide asserting that suicidal ideation turns into intention to act due to feelings of entrapment triggered by inescapable defeat and humiliation appraisals [104]. It has three phases (premotivational, motivational, and volitional), the transition between them being facilitated or obstructed by relevant threat-to-self, motivational, and volitional moderators, respectively [104]. O'Connor's is the most recently developed theory of suicide, and the research testing its major components is still in its incipient stages.

1.3.2 Conceptual framework – Interpersonal Theory of Suicide

To define the purpose of our research and to get a better understanding of how risk factors for suicidal ideation and behaviour interact, we use a conceptual framework based on the Interpersonal Theory of Suicide [21,100]. Drawing on previous research, and claiming to address the majority of inconsistencies present in other theories of suicide [21], the Interpersonal Theory of Suicide has already garnered empirical support among both civilian [60,106-112] and military populations [113-119], thus providing a useful ideation-to-action framework for the conceptualisation of suicide risk [102].

According to the theory, three distinct variables must be present for an individual to die by suicide: thwarted belongingness, perceived burdensomeness, and acquired capability. The first two are primarily related to suicidal desire. ***Thwarted belongingness*** is defined as the perception that one lacks important, strongly desired connections and meaningful relationships with others. According to the theory, deleterious measures of social connectedness and support (e.g. lack of perceived social support, self-reported loneliness, living alone and having low levels of social interaction, loss of a spouse through death or divorce, religious affiliation, marriage, number of children and friends) are indicators of thwarted belongingness that enhance the risk for passive suicide ideation [21,100]. The belief that one is a burden on others and a liability to others' well-being encompasses ***perceived burdensomeness***. Stressful life experiences (e.g. unemployment, financial difficulties, loss through death or separation, chronic pain, physical illness, childhood abuse, military sexual trauma, and combat exposure) are considered indicators of perceived burdensomeness, and therefore predictors of passive suicide ideation. According to the theory, thwarted belongingness and perceived burdensomeness are related but distinct constructs [106] and can predict passive suicidal ideation ("I am better off dead") individually [120] or in combination [107,121], while further interaction between these two constructs and ***hopelessness*** about the future ("This will never change") is necessary for active suicidal ideation ("I want to kill myself") to emerge [21]. However, suicide desire alone is not sufficient to drive an individual to attempt suicide. To result in lethal suicidal behavior, the desire to die must be accompanied by an ***acquired capability for suicide***, defined as the degree to which one is able to enact a lethal suicide attempt [21,100]. The capability for suicide is acquired over time through repeated exposure to experiences inducing fear and pain that result in habituation and therefore, a sense of lowered fear of death and an increased tolerance for physical pain. Previous suicide attempts are considered among the most potent behaviors with regard to acquiring the capability for suicide, as they are engaging the individual in habituation with regard to fear and pain involved in self-injury [21,100]. Combat exposure and childhood maltreatment are other, less potent but important, indicators of acquired capability for suicide, as they supposedly lead through repetition towards an increased tolerance to pain and fearlessness of death [21,100]. Moreover, the easy access to the weaponry in the military provide the opportunity for the military personnel to engage into habituation to the fearsome aspects of the use of the most lethal means, such as firearms, and thus acquire a sense of lowered fear of death [21]. Not surprisingly, firearms was the most common method of suicide in the CAF in 2015, accounting for a third of all suicides [18].

Thus, according to Interpersonal Theory of Suicide, thwarted belongingness, perceived burdensomeness, and acquired capability are not sufficient by themselves to lead to the development of suicide plans and attempts. It is only the combined presence of all these three relatively rare constructs that lead to the rarer act of lethal self-harm, making the theory consistent with the mismatch between sizable rates of suicidal thoughts and plans versus low base rate of suicidal attempts in the population [21]. According to Interpersonal Theory of Suicide, mental disorders increase the suicide risk *indirectly,* by conferring an elevated risk for experiencing lower levels of belonging, perceived burdensomeness, and painful and provocative experiences [21,100].

Figure 1.1 depicts the main causal and interactive constructs of Interpersonal Theory of Suicide, alongside the empirically-supported risk factors for suicidal behaviour that represent the observable indicators of these constructs. The causal process starts with thwarted belongingness and perceived burdensomeness and ends with near or completed attempted suicide. Thus, the lowest risk for suicide is towards the left side of the figure, comprising those individuals with passive suicidal ideation, with the risk increasing incrementally towards intent and attempt, represented towards the right side of the figure. According to the theory, suicidal ideation results

from the fewest number of co-occurring risk factors, suicide intent from a greater number, while suicide attempts or completions are the consequence of the co-occurrence of the greatest number of risk factors [21].

1.4 Suicide prevention and research in the Canadian context

During the past decade, the CAF strengthened its mental health system to better serve the needs of its members, among whom more than 40,000 have been deployed in support of a difficult military operation in Afghanistan. Notwithstanding the relative stability of CAF suicide rates over the past two decades, the most recent statistics showed a trend towards an increased risk for suicide for deployed CAF members compared to those who never deployed [17,18], suggesting that both mental health services renewal and deployment of many personnel on a difficult military operation may have changed the dynamics of mental disorders and suicidal behavior in CAF.

An up-to-date assessment of the mental health status of CAF personnel was therefore judged to be essential to gather an accurate picture of the mental health of CAF personnel in general and suicide behavior in particular. For this reason, the CAF commissioned Statistics Canada to undertake a large-scale cross-sectional mental health survey in 2013, the Canadian Forces Mental Health Survey (CFMHS). Its primary objectives were 1) to evaluate the impact of the mission in Afghanistan on mental health status and functioning of currently-serving personnel, and 2) to assess the adequacy of the CAF's renewed mental health system in meeting the needs of its members [122]. This survey used similar data collection methodologies as the Canadian Community Health Survey-Mental Health (CCHS-MH), designed and conducted by Statistics Canada in 2012 [123], thus offering researchers a unique opportunity to explore civilian-military comparisons on mental health issues, including suicidal ideation, plans, and attempts.

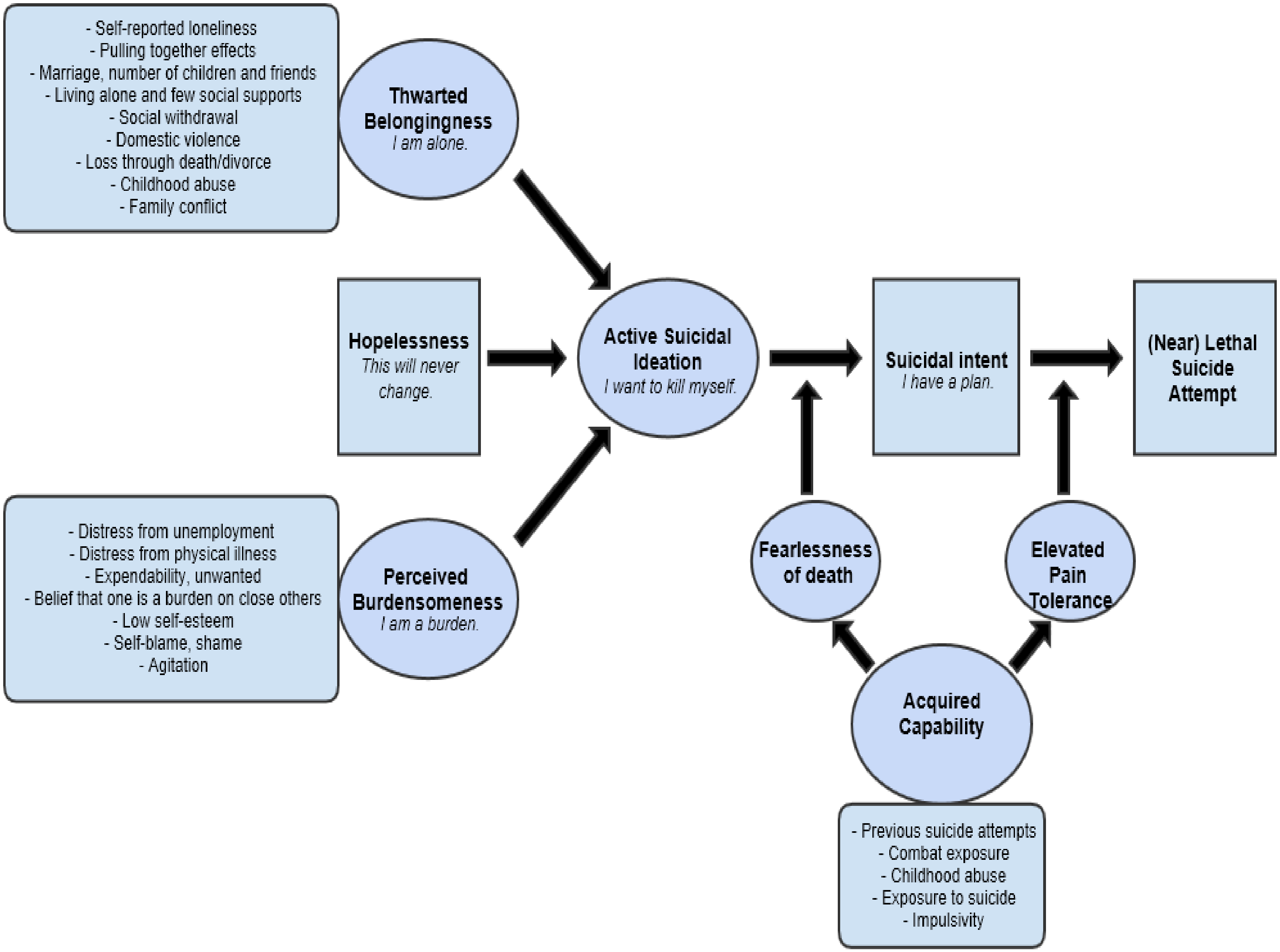


Figure 1.1. Pathways to (near) lethal suicide attempt according to the Interpersonal Theory of Suicide (adapted from [21])

Public interest notwithstanding, suicide prevention in the military is one of the top priorities of the Department of National Defense and the CAF, and military mental health has remained a topic of intense scientific interest during the past decade. The CAF's approach in building its own suicide prevention program was to include the most relevant components of the US Air Force's benchmark suicide prevention program, for which the most relevant evidence of efficacy exists, and to supplement it with the most effective evidence-based prevention strategies identified in the civilian literature [12,124]. Recently, a CAF Expert Panel on Suicide Prevention identified the need to reinforce the CAF's infrastructure to support quality improvement for its mental health programs and services as the primary opportunity for augmenting the CAF suicide prevention program [12].

At the federal level, the Public Health Agency of Canada has been tasked with producing a national strategy for suicide prevention in consultation with relevant federal departments, provincial and territorial governments and non-governmental organizations [92]. The Federal Framework for Suicide Prevention that has been recently released, though not a national strategy, provides guiding principles for joint action between the Government of Canada and all relevant sectors to help preventing suicide, by raising public awareness and reducing stigma associated with suicide, disseminating information and data to help prevent suicide, and promoting the use of research and evidence-based practices in suicide prevention [4].

1.5 Challenges of current suicide research

The worldwide recognition of suicide prevention as a public health priority has encouraged research in this area during the past decades, leading to the development of several prevention strategies with promising results in lowering the suicide rates. A recent meta-analysis cited restriction of access to lethal means, treatment of depression, ensuring chain of care, and school-based universal prevention as the strategies with the highest prospects for curbing the suicide rates if applied on a large scale [125,126].

Notwithstanding the increase in research and knowledge about suicide behaviour, the current preventive measures seem to do little in substantially improving the situation. Canadian vital statistics are disappointing: the mortality rate due to suicide did not decline over the past decade [5]. Stagnating rates has been observed for non-fatal suicide outcomes too. Large civilian and military mental health national surveys, such as CCHS-MH 2012 and CFMHS 2013, indicate that the past-year prevalence of suicidal ideation and suicide attempts did not change since 2002 [127]. Similar patterns have been observed in other nations [128]. Moreover, recent meta-analyses of the past 50 years of research on risk factors for suicide thoughts and behaviours concluded that predicting suicide behavior is no better than tossing a coin [101,129].

Every identified risk factor is a potential avenue towards determining risk groups in the population and the type of intervention required, being a key component of any effective prevention strategy [130]. However, on its own, each risk factor provides limited predictive power in terms of magnitude of effect and specificity [101,131].

To get out of this half-a-century deadlock, various researchers in the field called for adopting new methodologies, such as machine learning algorithms that allow for combining risk factors in ways that may increase the predictive accuracy of suicide behaviour [36,101,129,131].

Unlike traditional statistical techniques, such as logistic regression, machine learning methods are capable of accommodating hundreds of predictors and highly complex interactions between them and show a great potential not only for improving the predictive accuracy of suicidal behaviour, but to do it with potentially smaller sets of interacting predictors [132-137].

1.6 Predicting suicidal behaviour: Are we any closer?

Every day clinicians around the world are tasked to determine whether their patients will engage in suicide behavior. Judging by the plethora of risk factors that have been identified during past half a century of suicide research, we may be tempted to assert that the empirical evidence accumulated so far has been able to provide clinicians with all the necessary tools for such a daunting task. However, recent reviews of past 50 years of suicide research have soberingly concluded that, *within the narrow limits of existing literature*, we are no better than random coin-tossing in predicting suicide behaviour [101,129,138,139]. This is a rather bold statement, but makes sense when large community studies find that oft-cited risk factors, such as mental disorders, are in fact weak correlates of suicide attempts [47], or that very few of these factors discriminate well between ideators and attempters [24]. One such meta-analytic review found that currently known biological risk factors are poor predictors of future suicidal behaviors [138]. Another meta-analysis, focusing on anxiety and its disorders revealed that, when studied in isolation, anxiety constructs are weak predictors of suicidal behaviors over long follow-up periods [139]. A third meta-analysis that looked at history of self-injurious thoughts and behaviors concluded that its potential of predicting future suicide behavior was near chance levels [129]. Finally, one extended meta-analytical review that looked at broad risk and protective factor categories found that existing risk factors are weak and inaccurate predictors of both suicide attempts and suicide deaths [101].

Another overarching conclusion that stemmed from this review were that even when an association exists, it is so weak that it does not translate into useful clinical information. This lack of predictive accuracy can be traced back to at least two sources: (1) a lack of specificity of known risk factors corroborated with a lack of distinction between feeling suicidal and acting on suicidal thoughts in most of the available theoretical underpinnings of suicidal behavior [102], and (2) a focus on studying risk factors in isolation and/or static fashion, rather than in conjunction with other factors [101].

However, this does not imply that known risk factors are irrelevant to suicide behaviour, but that the present results are contingent on the constraints of the methodologies used to study these predictors. There is a growing need for a major rethinking of the methods with which the researchers tackle this phenomenon. Currently used analytical techniques should be shifted towards, or at least supplemented with, novel machine learning methodologies that not only can study hundreds of risk factors concurrently, but are able to produce risk algorithms with better predictive performance that conventional methods [36,101,129,131].

1.7 Aim and objectives

Building on the literature on risk factors for suicide behavior and the conceptual framework delineated by Interpersonal Theory of Suicide, the present research uses recursive partitioning algorithms to investigate cross-sectional predictors of past-year suicidal ideation, suicide plans, and attempts in two population-based samples of civilians and active military personnel. Both mental health surveys used similar methodological approaches to collect the data, offering a unique opportunity to explore civilian-military comparisons, while trying to shed light on whether the risk factors for suicidal behaviour in the Canadian civilian population can be generalized to military personnel and vice-versa. One particular challenge for the present research is the fact that many of the predictors considered here are potentially derived from highly correlated constructs [140-142]. Without addressing this thorny issue, it will be difficult to determine which predictors discriminate best between individuals having experienced past-year suicidal behaviour and individuals who have not.

The objective of this research is three-fold:

(1) To identify factors discriminating best between individuals having past-year suicidal ideation, suicide plans and attempts, and individuals who have not, in Canadian civilians and military personnel;

(2) To synthesize this information by generating risk algorithms based on subsets of most influential predictors with the ultimate goal of maximizing the predictive accuracy of suicide ideation, plans, and attempts;

(3) To determine whether the risk algorithms of past-year suicidal behavior differ in military personnel when compared with the Canadian civilians, and whether they can be generalized within and between samples.

First, we will focus on identifying most relevant correlates of suicidal ideation, suicide plans, and attempts in both samples, based on commonly measured features (Objective 1). This will give us an up-to-date picture of the most discriminative risk factors for past-year suicide ideation, plans, and attempts in the Canadian civilians and military personnel.

Following the tenets of Interpersonal Theory of Suicide, we hypothesize that passive suicide ideation, hopelessness, as well as different constructs of thwarted belongingness (living alone, low sense of belonging to local community, low scores on social provision scale items), and perceived burdensomeness (lost self-confidence, worthlessness, guilt) interact to predict active suicide ideation, and consequently will rank high among the cross-sectional predictors considered here (Hypothesis 1a). Furthermore, active suicide ideation will predict suicide plans and attempts via interaction with constructs of acquired capability of suicide, such as childhood adverse experiences.

We also hypothesize that, while models on each of the three aspects of the suicide behavior will overlap on several risk factors, unique covariates will differentiate them (Hypothesis 1b). For example, we expect that adverse childhood experiences, which can be construed as both a measure of thwarted belongingness or acquired capability of suicide will be associated with all three aspects of suicide behavior [21], while depression-related constructs, through their indirect effects on thwarted belongingness, perceived burdensomeness, and will be associated more strongly with suicide ideation than suicide attempts [20].

We will take advantage of the recursive partitioning methodology to select parsimonious subsets of most influential predictors, while accounting for the correlation between them, and generate models that maximize predictive accuracy, with similar performance as more complex ones (Objective 2).

We hypothesize that the majority of the covariates considered here are not relevant to the outcomes, and that only a handful of most influential predictors will be selected (Hypothesis 2a). According to the broad categories of risk factors examined by Franklin et al. [101], we would expect that hopelessness, depression-related, and anxiety-related constructs, and any child abuse will predict suicide ideation, while psychiatric hospitalization will predict suicide attempts.

In terms of predictive performance by outcome type, we hypothesize that models for suicidal ideation will have a lower accuracy than the ones for suicide attempts, because more severe outcomes are usually associated with more discriminative predictors (Hypothesis 2b). However, this may not be the case, because models for suicide ideation are fitted on more cases than those for suicide attempts, and therefore may yield similar or even better performance statistics.

The third objective is to determine whether the correlates of suicidal behavior differ in CAF personnel when compared with Canadian civilians (Objective 3). This is of paramount importance for the evaluation of the CAF suicide prevention program as it has recently incorporated nearly all of the suicide prevention strategies most consistently identified in the civilian scientific literature [12]. Both samples will be restricted to commonly measured variables, thus enabling direct comparisons between the groups by using cross-validation techniques following the creation of models using binary recursive partitioning.

The active military personnel are generally healthier relative to civilians in the same age group range, due to the rigorous selection procedures required for entering the military service and an ongoing demand for maintaining a standard of well-being among military personnel [143]. They also differ from their civilian counterparts in many essential aspects related to military culture and work environment. Unlike civilians, military personnel benefit from an increased sense of social connectedness that buffers the risk for suicide, due to a strong sense of community and camaraderie among peers who share the same professional ethic, ethos, and value system [20,38]. These differences will likely play an important role in establishing whether the relation between each risk factor and suicidal behavior differ when comparing military personnel with the civilian population.

Since military personnel have a unique socio-demographic composition, and are a distinct part of the labour force, we expect the models for suicidal behaviour to vary accordingly in the two samples (Hypothesis 3a). In other words, due to the distinctive elements of military life, we hypothesize that some of the risk factors strongly associated with suicide behavior in civilians, such as unemployment and lack of social integration and support, will play a lesser role in predicting suicide behavior in the military.

Due to military-civilian differences, predictive accuracy should vary by sample type. We expect that an algorithm applied on a civilian sample will perform less accurate in predicting any of the three outcomes in military personnel and vice-versa (Hypothesis 3b). In other words, we hypothesize that the predictive accuracy between samples will be lower than the one within samples [144].

2. Machine learning overview

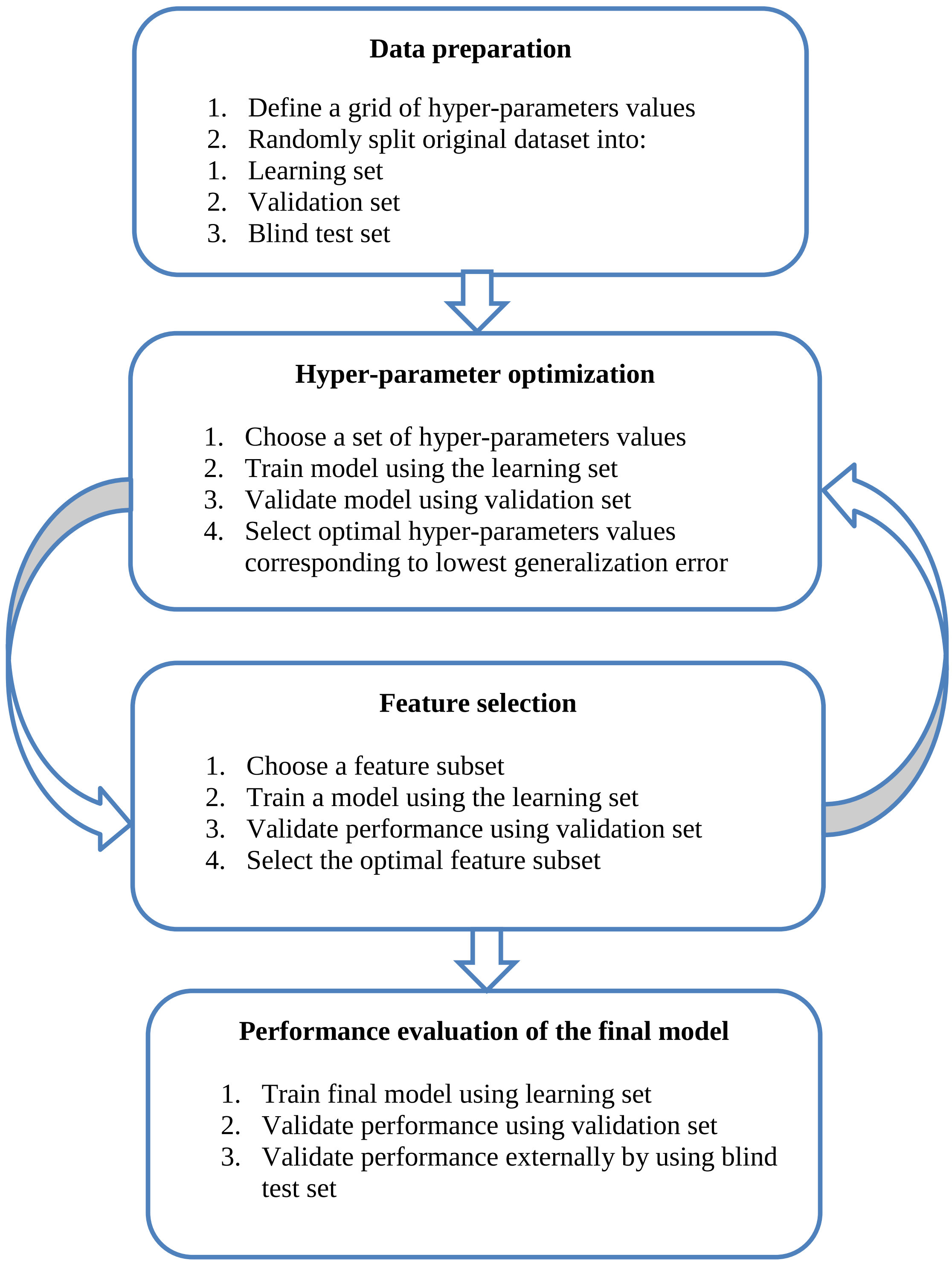
Often in medical research, the goal is to find a decision rule for predicting the status of a subject, based on a potentially large dataset of previous cases. The interest lies in predicting the next specific outcome (i.e. experiencing past-year suicidal behavior outcomes), but is often supplemented by a quest for identifying which factors, alone and in interactions, discriminate best between individuals having the outcome, and those who have not [145-147]. Used in conjunction, accurate classification and understanding the predictive structure of the data can help in identifying particular groups at risk as well as specific targets for risk factor reduction. Machine learning methods are one of the most efficient means of accomplishing both tasks.

2.1 Definition

Machine learning is an automated learning process based on *inductive inference* that provides algorithms able to explain or predict a phenomenon based on past observations [146,148]. In the present study, the phenomenon is a *discrete binary response* variable (i.e. whether a subject has a history of past-year suicide ideation, plans and attempts), while the past observations are referred to as the set of *covariates* or *predictors* which were commonly measured in the Canadian general population and CAF personnel samples. Given this set of common predictors and their corresponding *classes* or *labels* for the binary response variable, the objective is to find a systematic way of predicting a subject’s class conditional on the observed values of the covariates (Figure 2.1).

The learning method through which a relationship between the response variable and the set of covariates is unveiled and subsequent predictions are eventually generated is referred to as learning by example or *supervised learning* because the algorithm has access to the classes of the response variable [148]. The aforementioned objective can be also defined as the supervised learning task of producing a *model* or *classifier* able to approximate as closest as possible the labels of the subject’s set of covariates. The nature of the response variable determines the type of classifier, which in this case is a discrete binary classifier. The task is performed in two stages: *model training* and *model validation*, and its execution is controlled by algorithm-specific tuning attributes, called hyper-parameters [149].

Figure 2.1. General framework for developing a machine learning algorithm



2.2 Model training

During *model training*, the algorithm uses a pre-defined set or *grid* of values for its *hyper-parameters* to perform a classification task to determine model-specific attributes called *parameters* from a labeled set of covariates contained in a *learning* or *training set* [149]. The classification task is performed in a fit method in which the values of the covariates are mapped onto the labels of the binary response variable.

2.3 Model validation

It is of paramount importance for any type of predictive algorithm to be able to generalize well to new datasets from the same domain. Generalizing on the same observations that have been used in the process of generating the algorithm yields overly optimistic results, because it reflects how well the method reproduces known labels, rather than extrapolating on unseen ones [145]. The evaluation of a classifier predictive behavior is referred to as *model validation* and it is assessed by using covariate values from a set of subjects whose outcome labels are not known to the classifier, called *validation* or *testing set* [150]. A predictive method which uses model-specific parameters obtained during the learning task generates predictions as output labels [149]. Furthermore, the learning algorithm can modify the relationship between the response and the covariates based on the differences between the original and the predicted values of the response [151]. Finally, predictive accuracy is assessed by the degree of disagreement between the true class of the subject in the validation set and the one assigned by the classifier, which is referred to as the *prediction* or *generalization error*. The objective of most classifiers is to minimize this error [151,152]. For discrete binary classifiers, the prediction error is the *misclassification rate*, which is the fraction of subjects incorrectly classified. Conversely, the *predictive accuracy* of a classifier is the fraction of subjects correctly classified. The misclassification risk is modeled as a zero-one loss, in which the classifier incurs a loss of 0 for every correctly classified subject, and a loss of 1 otherwise [150].

2.4 Model selection

Algorithms come with default hyper-parameters setting values. However, relying solely on these values does not guarantee the lowest prediction error. The predictive accuracy of a classifier is conditional on both the size of the learning set [145,150] and the user-specified hyper-parameters of the algorithm [153]. In other words, depending on how the original data is split into a learning and a validation set, as well as on the choice of the values of the hyper-parameters, the algorithm can infer many classifiers from the data, each with its own predictive accuracy. Choosing the optimal model from a set of candidate classifiers is referred to as *model selection*, which involves optimizing the algorithm's hyper-parameters from a *tuning grid* of candidate values with the objective of selecting those settings that yield the best predictive performance [150,153,154].

2.5 Performance measurement metrics

A comprehensive picture of model performance cannot be derived solely on misclassification rate. Additional measures are needed, such as statistics derived from the confusion matrix, which is a visualization tool typically used in supervised learning (Figure 2.2), and the receiver operating characteristics (ROC) curves. We define them as follows:

1. *Accuracy* (precision) is the is the probability of correctly classifying a true (suicidal and non-suicidal) case. Its complement is the misclassification rate;

2. *Sensitivity* (recall, true positive rate) is the probability of correctly classifying a true suicidal case;
3. *Specificity* (true negative rate; TNR) is the probability of correctly classifying a true non-suicidal case;
4. *Precision* (positive predictive value; PPV) is the probability that an individual classified as suicidal is a true suicidal case;
5. *Negative predicted value* (NPV) is the probability that an individual classified as non-suicidal is a true non-suicidal case;
6. *ROC curves* is a graphical technique depicting relative trade-offs between true and false suicidal cases [155]. It is a plot of sensitivity against false positive rate (1-specificity);
7. *Area under the ROC curve (AUC)* is a reduction of a ROC curve to a single scalar value representing the expected performance of a classifier [155]. It is expressed as the probability that the classifier will rank a randomly chosen positive suicidal case higher than a randomly chosen non-suicidal case. A perfect classifier has an AUC of 1, while a random one has an AUC of 0.5.

Figure 2.2. Confusion matrix

	True suicidal	True non-suicidal
Classified suicidal	True positive=TP	False Positive=FP (Type I error)
Classified non-suicidal	False negative=FN (Type II error)	True negative=TN

Accuracy=(TP+TN)/(TP+TN+FP+FN)=(TP+TN)/n
Sensitivity=TP/(TP+FN)
Specificity= TN/(TN+FP)
Precision=TP/(TP+FP)
Negative predictive value=TN/(TN+FN)

2.6 Methods for estimating the prediction error

The methods used for estimating the prediction error of a classifier fall into two main categories: hold-out approach and methods based on resampling [150]. Their choice is usually driven by the size of the dataset at hand.

2.6.1 Hold-out method

When the original dataset is large, a sample completely independent of the learning set can be put aside. This is the *holdout method* (Figure 2.3) in which a training set is obtained by drawing *without replacement* $\mathbf{n}<\mathbf{N}$ subjects form the original dataset. The complement **N-n** is the testing set or the *out-of-bag (OOB)* observations, consisting of all subjects not drawn. The size of the testing set is usually fixed at 1/3 of the original dataset, but there are no theoretical justifications for this choice [145]. Other sample size ratios may be used as well. After generating the sets, the algorithm is trained on the learning set and a classifier obtained. The classifier is subsequently given the subjects from the testing set and an estimate of the prediction error is obtained. If the dataset is large enough, the testing set can be further split into two disjoint subsets. This is the ideal approach for obtaining the optimal values of the algorithm's hyper-parameters. Thus, the

Figure 2.3. Hold-out Sampling Method

2. Obtain the learning set by drawing *without replacement* a subset of subjects $\mathbf{n}<\mathbf{N}$ from the original dataset
3. Obtain the testing set by keeping track of all subjects not included in the learning set
4. Train the algorithm on the training set and generate a classifier
5. Test the classifier on the testing set and obtain an estimate of the prediction error

algorithm is presented with a *learning set* for training, a *validation set* for model selection, and a *blind testing set* for evaluating the performance of the best classifier [150]. The main advantage of holdout method is that the estimate of predictive error is independent of the learning set. However, by splitting the original dataset, the method restricts the number of observations used for training, which is a challenge when available samples are small [145]. The method is less feasible in these circumstances, because the algorithm will not train well if there are not enough subjects in the learning set, while too few cases in the testing set lead to suboptimal estimates of the prediction error [150]. On the other hand, if the entire dataset is used for training, it is not possible to assess the classifier's performance on new data, because all the observations have been included in the learning set. For small datasets, methods based on resampling are preferred [145].

2.6.2 Simple resampling

Simple resampling refers to methods is which every subject in the original dataset is used only once for validation [145,150]. One of the most utilized simple resampling methods is the ***k**-fold cross-validation* (CV) in which the original dataset is split initially into **k** disjoint testing sets or *folds* of roughly same size (Figure 2.4). Then the algorithm is trained on **k**-1 folds and tested on the **k**-th fold. The process is repeated **k** times and **k** estimates of prediction error are obtained. Finally, these estimates are averaged to obtain the classifier's mean estimate of prediction error. There is no formal rule for choosing **k**, but higher values will generally yield estimates with low variance and relatively low bias [153,156]. The most common used values for **k** in machine learning algorithms are 5 and 10 [150,153].

2.6.3 Multiple resampling

Multiple resampling refers to methods in which a subject is used several times during model validation. The estimates of the prediction error are generated either by performing simple resampling multiple times (*repeated k-fold CV*), or by drawing many samples from the original dataset (*bootstrapping, random subsampling*).

In repeated k-fold CV, a **k**-fold CV method is repeated **m** times and the mean of all **m·k** estimates of the prediction error is obtained (Figure 2.4). The most common used values for **m** in machine learning algorithms are 5 and 10.

Figure 2.4. k-fold Cross-Validation and Repeated k-fold Cross Validation Resampling Methods

k-fold Cross-Validation

1. Draw *without replacement* **k** disjoint same-size testing sets from the original dataset
2. Put aside the **k**-th testing set, and use the remaining **k**-1 testing sets to train the algorithm and generate a classifier
3. Test the classifier on the *k*th testing set and obtain an estimate of the prediction error
4. Repeat step 2 and 3 **k** times
5. Average all **k** estimates to obtain the mean estimate of prediction error

Repeated k-fold Cross Validation

1. Set up the number of iterations **m**
2. For each iteration **i≤m** DO
3. Draw without replacement **k** disjoint testing sets form the original dataset
4. Put aside one testing set at a time, and use the remaining **k**-1 sets to train the algorithm and generate a classifier

5. Test the classifier on the testing set and obtain an estimate of the prediction error
6. Repeat step 4 and 5 **k** times
7. Average all **m·k** mean estimates to obtain an overall mean estimate of the prediction error

The bootstrap sampling method consists of drawing *with replacement* multiple (**m**≥200) same-size resamples form the original dataset [150,156]. Each resample is a training set and its complement, including every case not drawn, constitutes a testing set. The algorithm is trained and tested as usual and an estimate of the predictive error is obtained. Finally, all the **m** estimates are averaged (Figure 2.5). Overall, around 63.2% of all distinct cases from the original dataset appear in the **m** training sets, while the remaining 36.8% fall into the **m** testing sets [150,156]. Thus, in bootstrapping, the algorithm is trained on roughly 2/3 of the data and tested on the remaining 1/3.

Random subsampling is an extension of the hold-out method. The procedure is repeated multiple (**m**) times and estimates of the prediction error are averaged (Figure 2.5).

Figure 2.5. Bootstrapping and random subsampling

Bootstrapping

1. Obtain the training set by drawing *with replacement* a set of **N** subjects from the original dataset
2. Obtain the testing set by keeping track of all subjects not included in the training set
3. Train the algorithm on the training set and generate a classifier
4. Test the classifier on the testing set and obtain an estimate of the prediction error
5. Repeat step 1 to 4 **m** times
6. Average the prediction error estimates obtained during the **m** iterations

Random subsampling

1. Obtain the training set by drawing *without replacement* a subset of **n**<**N** subjects from the original dataset
2. Obtain the testing set by keeping track of all subjects not included in the training set
3. Train the algorithm on the training set and generate a classifier
4. Test the classifier on the testing set and obtain an estimate of the prediction error
5. Repeat step 1 to 4 **m** times
6. Average the prediction error estimates obtained during the **m** iterations

The main advantage of multiple resampling over simple resampling lies in the stability of the estimates due to the multiple uses of the data. However, this comes with a cost that poses a real challenge when evaluating the performance of a classifier against others: the loss of independence of the data used for generating the algorithm [150,156]. Instead of being obtained on independent sets, each classifier is the result of training and testing the algorithm on overlapping resamples, increasing the probability of type I error, and biasing the performance estimates [150].

2.7 The class imbalance problem

One issue that plagues performance of most machine learning algorithms, regardless of the sampling method employed, is the sparsity of cases in the data. This is referred to as the *class imbalance problem*, which has a strong impact on the predictive accuracy of a classifier [154,156-158].

A dataset is imbalanced if the number of observations belonging to one class exceeds that in the other class. Such datasets are common in the medical field, in which the number of subjects experiencing the outcome is much lower than those without it. Suicide and non-fatal suicide behavior are rare phenomena. It comes as no surprise that both CCHS-MH and CFMHS datasets are heavily imbalanced with regards to past-year suicidal ideation, plans, and attempts. Indeed, the prevalence range for any of the three outcomes does not exceed 5% in either dataset [8].

The critical issue with employing classification methods on class-imbalaced data is that they often produce classifiers that do not accurately predict the minority class. The prediction is biased toward the majority class. A predictive model can still get a very high accuracy by simply predicting those individuals without suicidal behaviour, because the algorithm, being developed to decrease the generalization error, and expecting balanced class distributions, will prefer to classify a subject as non-suicidal [158]. Specificity increases dramatically, while sensitivity drops to very small values [154,156]. However, the most relevant class should be the minority one, because the risk of misclassifying a case as non-suicidal has dire consequences for the individual and their family, friends and community: death for the former and profound trauma for the latter.

One way to handle the class-imbalance problem is to balance the classes at the data level by over-sampling the minority class (*up-sampling*), under-sampling the majority class (*down-sampling*), or creating synthetic minority class instances [159]. Several studies comparing these balancing methods on different machine learning algorithms found a net advantage of under-sampling [158,160,161].

While most researchers share similar views regarding balancing data *during* training, not all of them agree on performing the same procedure on the testing set. Some suggest using balanced datasets in both instances [156], while others advice against using balanced testing sets, arguing that the validation should be done on a sample that closely resembles the population from which it was drawn [153]. Moreover, balancing the data *before* training results in overly optimistic estimates. *Subsampling during resampling* avoids such outcomes, because it ensures that data points are appropriately split, and the resampling scheme captures the variation generated by subsampling [153].

2.8 Assessing performance of multiple classifiers

Finding the predictive accuracy of a classifier is not sufficient to evaluate its performance, as it frequently needs to be assessed against other classifiers to ensure that the observed difference in prediction error is not subject to *random variation*. Paired t-test is the statistical significance method of choice for assessing the differences between two prediction error estimates. In a paired *t*-test, the independence of the observations can be reasonably assumed if the values from which they are derived have been drawn without replacement from the original dataset. However, this is not the case when multiple trials are performed to obtain the error estimates. In these situations, the training sets overlap, leading to prediction errors that are not independent. This results in an underestimation of the variance across trials and a higher probability of type I error [150,156]. Thus, the paired *t*-test needs to be corrected to adjust for the correlation between prediction errors. Depending on the resampling method involved, several corrected versions of the *t*-statistic have been proposed [162,163]. Assessing multiple classifiers finds its utility during model selection, when a classifier with similar predictive accuracy as the best model, but with lower values for the user-specified hyper-parameters, is chosen. Classifiers are generated through computationally intensive methods, and running the procedure on higher hyper-parameters values significantly increases the computational burden. Thus, relying on a less complex model is desirable when computer resources are limited.

2.9 Variable selection

If the set of candidate predictors is large, it is essential to retain only the predictors that best describe the phenomenon under study. Keeping non-informative or highly correlated predictors adds uncertainty to the prediction error estimates, reducing the performance of the classifier [153]. Discarding redundant variables allows the model to reduce its complexity, with little loss of information, while removing highly correlated ones lowers the model variance and prevents over-fitting, leading to a more parsimonious and interpretable classifier, with relatively low alteration in predictive performance [153]. This is especially relevant to psychiatric epidemiology, because any group of symptoms is unlikely to be specific to a single mental disorder and self-report instruments that assess these symptoms are often highly correlated [140-142]. Therefore, mitigating the effects of both non-informative and highly correlated predictors is of outmost importance when building a classifier.

Depending on the type of algorithm used, model selection may involve a subsequent *feature or variable selection* step, which is a built-in process of determining the minimum set of covariates that ensures the lowest misclassification error. This dimensionality reduction technique is an important aspect of machine learning process, in which redundant or *noise variables* are removed by analyzing the contribution of each covariate to the overall performance of the model. The optimal classifier is no longer the fined tuned model containing all the covariates, but the one retaining only a reduced set of relevant predictors whose misclassification rate is closest to the one of the full model [164,165]. The downside of this approach is an expected decrease in predictive accuracy once irrelevant variables have been removed [164,166]. In other words, there is a real possibility that no feature subset can ensure a lower error than the full set. However, various heuristic approaches may be applied to ensure that the misclassification error of the reduced model is within some agreed range of the smallest error [145,165].

Supervised variable selection provides the researcher with less complex, more interpretable, cost-efficient classifiers [153]. By reducing the dimensionality of the data, it lowers computational time, but adds another level of complexity in the process because it undermines the optimal values of the hyper-parameters obtained during training on the full set of predictors. In other words, there is no guarantee that the values obtained during model selection with the full set are equally optimal when using only a reduced set of variables [167]. Consequently, algorithm hyper-parameter optimization must be carried after every feature selection step (Figure 2.1), which increases overall computational time of the procedure.

It turns out that the predictive performance of a classifier can be successfully enhanced by including feature selection during model building [153]. Even though many machine learning algorithms can perform classification tasks, not all of them have feature selection capabilities. Neural networks, k-nearest neighbours, and support vector machines are examples of methods that do not provide insight regarding quantifying covariates impact on the response variable. Other algorithms, such as linear discriminant analysis, do have built-in feature selection tools, but require a considerable reduction of the number of available variables before deriving any model [147].

2.10 Decision trees

One method that gives insight into what variables are the most important with respect to the response variable is the decision tree. A decision tree is a classifier based on a completely data-driven paradigm that is expressed as a recursive partition of the space spanned by the predictor variables (*feature space*). When the response is binary, the decision tree takes the form of a binary classification tree.

2.10.1 Description of decision trees

A binary classification tree resembles an inverted tree, consisting of a root and a series of nodes connected though branches or splits. The root is a node with two downstream branches but without any upstream branch, while an internal node is a node that has one upstream and two downstream branches. In contrast, a terminal node or leaf is a node with an upstream branch and no outgoing branches. Any node with downstream branches (root or internal node) is also a parent node, while any node with upstream branches (internal or terminal node) is referred to as a child node.

2.10.2 Growing decision trees

A binary classification tree is build through binary recursive partitioning, which is a method that splits in a recursive manner the feature space into a set of distinct, rectangular regions or subgroups, in which the subjects with similar class labels for the outcome are grouped together [151,168]. First, the original dataset is split into training and validation samples. Starting with all observations in the training sample (root node) and a set of predictors, a binary classification tree is generated by examining all splitting variables and their possible cut-points, and repeatedly partitioning the training set into mutually exclusive subgroups (nodes) that are increasingly homogenous with respect to the outcome. An optimal split will send the subjects from a parent node into two child notes such that the homogeneity in the child nodes is improved over the parent node. For example, in a hypothetical tree of people at risk for suicide plans (Figure 2.6), suicidal ideation has the strongest association with the outcome and it is examined first. Each subject is moved to the first terminal node on the left if suicidal ideation is experienced, or to the first internal node on the right otherwise. Next, subjects from the first internal node are examined based on the next most important variable, depression. If subjects are positive for depression, they are sent to the next terminal node on the left, and to the right otherwise. The splitting continues until no internal node on the right can be split further.

2.10.3 Splitting criteria

To select the most discriminative variable and its corresponding best cut-point, the algorithm employs a split criterion, which is a measure of the strength of the association between the variable evaluated for splitting and the response [168]. Examples of split criteria are impurity reduction measures [145] or p-values of association tests [169]. The predictor variable with the highest impurity reduction (that is, the highest reduction of the node heterogeneity with respect to the outcome) or the lowest p-value is selected for splitting the training set into two internal nodes. These nodes are split further by the same or an alternative covariate, increasing the depth of the tree, until some stopping criterion is applied [145]. The main result of the partition is the lower impurity and variance of the descendant nodes when compared with the parent node [145,156]. The terminal nodes are not split further and are assigned a class label based on the most frequent outcome of the subjects within those nodes [145]. The splitting condition, group size and the proportion of subjects experiencing the outcome are also reported at each terminal node.

2.10.4 Stopping criteria

The recursive partitioning of the training set stops when some stopping criterion is fulfilled, such as when the subjects within a node are only of one class (pure node), or when the number of observations within a node reaches a pre-specified minimum or when a pre-specified minimum change in the impurity reduction is not reached by any of the variables evaluated for splitting [168].

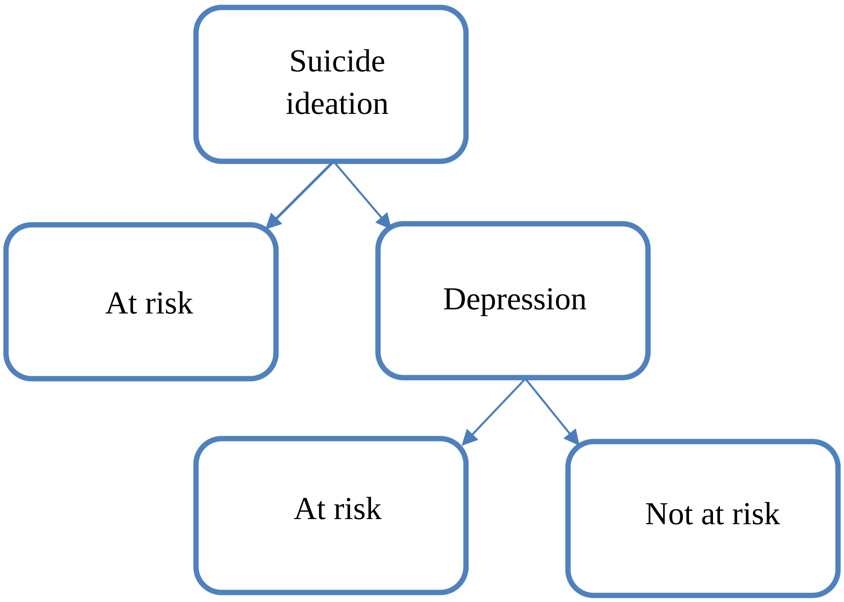


Figure 2.6. Hypothetical decision tree for suicide plans (adapted from [156])

2.10.5 Pruning

The tree built through binary recursive partitioning is large and overfits the data, because it captures both the underlying structure of the training set and the noise due to random sampling [168]. Therefore, its predictive accuracy on new independent samples will be poor, because it has too much variance due to its complexity [151]. To avoid overfitting, the tree is pruned using cost-complexity pruning, which is the process of removing those branches that do not add to the predictive accuracy of the tree in the validation sample [145]. How important a branch is a function of the misclassification rates and the number of nodes between the full-grown tree and the nested tree. The result of pruning is a set of subtrees, which are less complex trees nested within the original full-grown tree. The optimal tree is the one whose misclassification rate is within one standard error of the tree with the smallest absolute error [145].

2.10.6 Evaluating decision trees

Once the predictive factors have been identified, the validation samples will then be used to assess the accuracy of the prediction models obtained from the training samples. New subjects are classified by navigating them through branches from the root of the tree down to a terminal node and recording the class prediction of the outcome.

2.10.7 Advantages of single decision trees

A decision tree has several major advantages over the traditional statistical methods, the most important being that it is able to uncover nonlinear relationships and higher-order interactions between predictors that may be difficult or impossible to detect using traditional multivariate techniques [170]. While it assumes independent and identically distributed observations, a decision tree does not impose any assumptions regarding the underlying distribution of values of the outcome with respect to independent variables [145,153]. It allows missing data, it is robust to outliers and it is invariant to monotone transformation of variables [145]. It can handle various types of categorical predictors, as well as highly skewed or multimodal numerical variables without the need to pre-process them [145,153], and it is highly interpretable and relatively easy to implement [151,153]. In addition, a decision tree has a built-in feature selection process, which provides insights into what covariates are the most important with respect to the response variable [153].

2.10.8 Variable selection bias

The impurity reduction measure used in the original binary classification trees [145] is based on the Gini index which measures the divergence in probability distributions of outcome values within the descendant nodes [148]. It shows how frequently a randomly chosen subject from the parent node would be incorrectly classified if it was randomly classified according to the class distribution in the child node. The best cut-point of the predictor variable chosen for splitting is the one for which the decrease in the Gini index (impurity reduction) is the highest among all the predictors and their corresponding cut-points evaluated at that given node [156]. The impurity reduction is measured as the difference between the impurity in the parent node and the average impurity in the child nodes [168]. Maximizing the splitting criterion over all possible cut-points simultaneously leads to bias towards favoring multi-levele categorical variables [145,171,172], or predictors with missing values [173].

2.10.9 Disadvantages of using Gini index

The original binary classification trees suffer from selection bias that leads to trees with misleading structure. Several studies using simulated and real-world data showed that predictors with more categories are favoured over those with fewer categories when Gini index is used in the split process [172,173]. In other words, there is a high probability that the algorithm will select a noise variable over an informative one during the top splits, just because the former has more categories than the later. Indeed, a noise variable with k categories has $2^{k-1}-1$ potential cut-points to choose from, and is therefore more likely to be selected for an optimal split by chance alone and to sit higher up on the tree’s hierarchical structure than a binary informative variable [172] To make the matter worse, it has been shown that the selection bias is not restricted to multi-level categorical predictors, but to continuous ones and to predictors with missing values [169].

2.10.10 Conditional inference trees

Conditional inference trees are a type of single decision trees that were developed to mitigate the main flaw of original classification and regression trees, that is, the selection bias due to variables with many possible cut-points or missing values. The key differences between this type of decision trees and the classical classification and regression trees ones are: (1) the separation of variable selection and splitting procedures and (2) the use of a different split criterion, which is not based on the reduction of impurity within the nodes, but on p-values of association tests [169].

Conditional inference trees use unbiased recursive partitioning within a permutation test framework to select the node cut-point, based on the strength of the association between the covariate and the outcome ($\chi 2$ statistic). It involves three main steps [169]:

1. Starting with the whole sample, the algorithm evaluates the global null hypothesis of independence between any of the covariates and the outcome. The procedure stops if the global null hypothesis cannot be rejected. If it can be rejected (i.e. at least one covariate is associated with the outcome), the algorithm selects the variable with the strongest association with the response (smallest p-value, corrected for multiple-testing). This step alone ensures an unbiased variable selection, by selecting only the informative variables, and prevents over-fitting the tree (pruning is no longer needed);
2. The algorithm moves to finding the optimal binary split in the selected covariate, by means of $\chi 2$ statistic to maximize the discrepancy between the newly formed subgroups;

3. Previous steps are repeated until the global null hypothesis cannot be rejected at a given level of α (i.e. when no more statistically significant predictors are associated with the outcome) or when a minimum number of subjects per node is reached.

2.10.11 Advantages of conditional inference trees

The main advantages of conditional inference trees over the original decision trees are the mitigation of the selection bias by separating the variable selection and the splitting process, along with their flexibility of accommodating all types of predictor and outcome variables. Moreover, the pre-pruning strategy based on p-values is used to protect locally against splitting on variables that are unrelated to the response at a significance level α [169].

2.10.12 Disadvantages of single decision trees

The main flaw of single decision trees is their high variability [145,151,168,174]. This is an inherent consequence of applying binary recursive partitioning to a dataset in which random variability is present. Slight changes in the learning set yield unstable trees, because a different split variable may be chosen at the root of the tree [153,168]. This can be demonstrated by drawing several bootstrap samples from the learning set and fitting a decision tree for each sample [153]. The resultant trees will likely have different structures, impairing their interpretability and predictive accuracy.

Several solutions have been proposed to overcome the main disadvantages of decision trees. Unbiased classification algorithms have been proposed to overcome the selection bias [169,171], while new ensemble methods have been developed to reduce the high variability of decision trees [174,175].

3. Ensemble methods

Both CCHS-MH and CFMHS cross-sectional surveys contain hundreds of commonly assessed variables, many of them having an unknown relationship with any of the three outcomes, alone and in combination with other variables. Using parametric methods is unfeasible when dealing with a large feature space, as the modelling of a staggering number of high-order interactions is not possible. Even single decision trees are not very useful. Due to their high variability, their predictive accuracy is marginally above chance levels [175]. One efficient way to reduce the high variability of single decision trees is to use ensemble methodology that exploits trees' main weakness to make better predictions. The main idea of ensemble methods is to combine a set of many single decision trees, each of which solving the same classification task, to generate an aggregated model with better predictive performance.

3.1 Bagging

Bootstrap aggregation or bagging is an ensemble method in which a set of decision trees is generated by bootstrap sampling and their predictions are averaged [174]. The main user-specified hyperparameter in bagging is the number of trees to be grown (ntree) and can be optimized for better predictive accuracy. A researcher facing a binary classification problem will construct a set of ntree binary classification trees using ntree same-size bootstrap samples (drawn with replacement) from the learning set. Unlike single decision trees, the trees in bagging are grown to the largest extent possible and left unpruned. Pruning is not necessary, because deep trees have low bias and aggregation reduces variance [176]. Being maximal, each single bagged tree has low bias but still shares the high variability of the classical pruned tree. However, by averaging their predictions, their overall variability is reduced [151,174,175]. Averaging is done through majority voting which is done by recording the most commonly predicted class among all bagged trees. The size of the set (ntree) can be later increased until there is no further gain in predictive accuracy.

Lower variability translates into substantial gain in predictive accuracy over a single decision tree, which has been demonstrated by several theoretical, simulation and empirical studies [174,176-179]. However, the trees in the ensemble are correlated, preventing the method from further improving the predictive accuracy [151,153]. They have similar structures around the root, because the same strong candidate predictor variables are selected during the first splits in each tree. This is an inherent consequence of considering all the predictors at every split for every tree in the ensemble. Finding ways to de-correlate the trees was the next logical step that led to the development of low bias, low variance methods called random forests.

3.2 Random forests

Random forests [175] are machine learning algorithms for classification and regression that belong to the family of nonparametric ensemble method. They were developed to improve the variance reduction of bagging by lessening the correlation between trees through random selection of predictors [151,175]. In fact, bagging can be considered as a special case of random forests, in which the number of selected variables at each split is maximal.

3.2.1 Main characteristics of random forests

In his seminal work on random forests, Breiman demonstrated that random forests predictive accuracy is contingent on the accuracy of each individual tree and its correlations with other trees. He further showed that the upper bound of the generalization error can be lowered when averaging less correlated trees [175]. By aggregating maximal trees with relatively low bias, the method ensures the same low variance observed in bagging, which is further improved when not all the candidate predictors are made available at each split. Adding random feature selection to the tree-growing process generates more diverse and therefore less correlated trees, with the benefit of increasing the predictive performance of the ensemble [153,156,168]. Additionally, every time when strong predictors are not chosen during the random feature selection allows for weak candidate predictor variables to enter the ensemble, and their interaction effects to be evaluated [168]. This is an advantage over both single decision trees or bagging, where less competitive predictors are always outplayed by the stronger ones and their interactions effects are never given the chance to be discovered.

Random forest methodology has many characteristics that makes it suitable for a whole variety of classification problems: (1) it handles large sets of predictor variables with good predictive performance; (2) can be used in situations in which there are more variables than observations; (3) does not require any model specification beforehand; (4) supports highly-correlated variables; (5) identifies and incorporates high-order interactions among predictors; (6) does not overfit; (6) captures non-linear associations with the outcome; (7) ranks variables according to their ability of predicting the outcome [154,164]. Moreover, several empirical studies suggest that random forest methodology is not only capable of accurate predictions in a variety of contexts, but consistently outperforms other machine learning algorithms [180-183].

3.2.2 Growing random forests

Random forests methodology (Figure 3.1) consists of a collection of binary decision trees generated by randomly subsampling from the learning set. It differs from bagging in the way that introduces an extra layer of randomization in the tree-growing process by evaluating each optimal split on a random sample of predictor variables, instead of the full set [168,175]. This ensures that no single predictor variable dominates over others during the tree growing process, resulting in more diverse trees and better predictive performance than in bagging. The main user-specified hyperparameters in random forests are the number of variables to be randomly chosen at each split *mtry* and the number of trees to be grown *ntree*. In classification problems, the default value of *mtry* is the square root of the number of variables [164]. However, both hyperparameters can be optimized for improving the predictive accuracy of the ensemble. The optimization is performed using a tuning grid, in which different combinations of hyperparameters' values are chosen before starting the training. As in bagging, the trees in random forests are grown deeply through binary recursive partitioning and left unpruned. However, tree-based stopping criteria can be used to determine when to stop growing each tree.

3.2.3 Predicting with random forests

Random forests predictions start with predicting from each single tree. After growing, a new subject with given values for the predictor variables is dropped down each tree in the forest and its class prediction is recorded. In other words, each tree casts a vote for the most frequent class predicted in each of its terminal nodes. Finally, the predictions from all the trees in the ensemble are averaged by majority voting: the class voted by more than half of the trees is the prediction of the ensemble [151,168,175].

Figure 3.1. Description of random forests algorithm

1. Select the number of trees to grow (ntree) and the number of predictors to be used at each split (*mtry*)
2. Generate a learning set by random subsampling with or without replacement and set aside the remaining data (OOB data)
3. Grow a single decision tree using the learning set by performing each split as follows:
a. Select a subset of *mtry* predictors by sampling without replacement from the original set of p predictors
b. Select the best predictor among the subset predictors and partition the sample
4. Use stopping criteria to determine when to abort tree growing, and do not prune
5. Repeat steps 2 to 4 *ntree* times

3.2.4 Interpretability of random forests

Compared to single decision trees, where the relevance of a predictor corresponds directly to its position in the tree structure, random forest methodology does not provide a similar straightforward structure for easy interpretation [172]. It helps increasing the prediction accuracy of a model at the expense of interpretability [151]. However, it produces alternative measures for variable interpretation that reflect predictor's complex relationship with the outcome [151,168]. These variable importance measures are an important feature of random forests and are discussed in detail below.

3.2.5 Automated features of random forests

Selection of a reduced set of most influential predictors out of the full set is achievable with random forests because they have two built-in features, such as the out-of-bag (OOB) error, and the measures of variable importance. Combining them ensures an elegant way to achieve this goal.

Out-of-bag error

Out-of-bag error is a built-in measure of prediction accuracy of random forests [175]. It is a product of the tree-growing process by random sampling through bootstrapping, in which approximately 36.8% of the data is not used to construct the tree. These observations constitute the OOB sample for the respective tree, and can be successfully used as an internal validation set. In other words, every observation from the OOB sample is dropped down each tree in the forest and the class prediction is recorded. In a classification context, the OOB error is the averaged misclassification rate over all the trees in the forest, and is considered relatively accurate when compared to the one obtained from a blind testing set [154]. This removes the need for setting aside a test set before training the random forest algorithm.

Variable importance measures

In any given large dataset, the majority of explanatory variables are non-informative with respect to the outcome. Influential predictors are few. It is therefore desirable for any machine learning algorithm to produce some measure of predictor ranking, based on the contribution of each variable in predicting the response [151]. Fortunately, ensemble methods have built-in measures of variable importance that are computed to assess the relevance of each covariate over all trees in the ensemble. This importance scores can be subsequently used for ranking purposes to select a subset of most relevant predictors, rendering them as potential screening tools in epidemiological research [184].

3.2.6 Permutation accuracy importance

In random forests, the most advanced variable importance measure is the permutation accuracy importance [172]. In its original form, the relevance of a variable is determined by the increase in misclassification rate after its relationship with the outcome is lost through random permutation [175]. First, each tree is grown and misclassification rate is determined on each OOB sample. Second, the values of the variable in the OOB sample are randomly permuted to destroy any association with the outcome, while the values of all the other variables remain unchanged. Third, the trees are grown again, and a second misclassification rate for each tree is obtained. Finally, the difference in misclassification rates before and after permutation, averaged over all trees, constitutes the variable importance measure (VIM). The bigger the difference, the more important the variable is, and the higher its ranking. In contrast, a difference around zero indicate that the variable has no predictive value [168]. VIMs can successfully identify predictors involved in complex interactions, that is, variables that predict the outcome only in association with other predictors [154]. These high-order interactions would likely have been missed in single decision trees, bagging, and even parametric methods [168].

3.2.7 Dealing with correlated variables

Correlation between predictors strongly impacts variable selection [185,186]. This is critical when the feature space contains only a handful of truly influential predictors among a large set of highly correlated variables. In such scenarios, it is preferable to detect the influential subset more frequently than a non-informative one, by accounting for within-predictor correlation [186].

When a non-informative variable is highly correlated with other predictors, the choice of which predictor to use for splitting is driven by random differences between them [153]. Moreover, the non-informative predictor receives a higher importance score than other non-informative, but uncorrelated predictors, making it undistinguishable from truly informative ones [186]. Overall, more predictors are selected than needed and their importance is biased. Simulation studies found that the original variable importance measures are biased towards correlated variables. The bias emanates from the tree growing process, and from computing importance scores by means of unconditional permutation scheme [185,186].

Correlation between predictors can be properly handled at the algorithmic level, by means of a conditional permutation scheme [185]. This approach differs from the unconditional one by permuting one predictor within strata of sets of predictors that are correlated with it [185]. It attenuates the correlation between predictors, providing less biased variable importance measures.

3.2.8 Variable selection methods

One way to improving prediction accuracy with random forests is to use less predictors when growing the trees. The correlations between the trees are increasing with the number of predictors used to construct the ensemble, and the lowest generalization error can be achieved when only a small set of predictors are used [175]. Finding a smaller subset with similar of better predictive accuracy than the one achieved by the entire feature set is a desirable goal for model interpretation.

Several variable selection methods have been proposed to uncover the most relevant predictor variables from a potentially large set of covariates [164,165,187]. Their core element is the concept of variable importance measure. Most successful methods use a stepwise approach to reduce the feature space. In a nutshell, starting with the full model, fractions of least informative variables are sequentially removed from the feature set, the algorithm refit, and predictive accuracy and variable importance measures computed. The process continues until no further variables can be rejected, and predictive accuracies are compared. The most parsimonious classifier, whose predictive accuracy is within a specified number of standard errors of the best performing one, is chosen for validation (Figure 3.2). How the fraction of least important variables is rejected at each loop is the most critical step in the variable selection process. One rather arbitrary, but quick way to obtain the smallest model is to remove 20% of the least-informative variables before refitting [164]. Another method, involves a preliminary elimination step, in which useless predictors are removed first, by setting a threshold for the standard deviation of least-informative variables, followed by building increasingly complex nested models, based on variables ranking [165]. This approach is slower, but it is entirely data-driven, and has been shown good performance when compared with other variable selection methods [187].

Figure 3.2. Generic steps for variable selection

1. Start with the most complex classifier
2. Assess predictive accuracy and compute variable importance measures
3. Remove a fraction of least informative variables
4. Build a new classifier using the remaining variables
5. Repeat step 3 and 4 until no further variables can be rejected
6. Choose the most parsimonious classifier, whose predictive accuracy is within a specified number of standard errors of the best performing one

3.2.9 Random forests tuning

There is no guarantee that training over the default hyper-parameters settings (*ntree*=500 and *mtry*=√p) will generate the best performing classifier. Tuning *mtry* and ntree via tuning grids is highly encouraged, as both hyper-parameters impact model performance, and stability of variable rankings [185]. It has been suggested that the number of trees should be increased with the number of covariates, so that each candidate predictor is given the opportunity to be selected in the ensemble [154]. Denser forests are also indicated for stabilizing the ranking of classifier's most influential variables [185]. The general recommendation is to employ several values of ntree, in ascending order, until the prediction error levels off [154], while preserving the computational feasibility of the method. The effects of *mtry* on variable selection are even more dramatic than for *ntree*. However, no formal method of tuning exists for *mtry*. One heuristic approach involves starting with 5 arbitrary values, corresponding to splitting the interval between 1 and the size of the feature set into 5 equally distanced subintervals [153].

3.2.10 Conditional inference forests

Standard random forests methodology has a major pitfall: by using standard single decision tree as their base, it carries tree bias towards multi-categorical predictors [172]. In other words, in every tree of the ensemble, variables with more categories are favored during the split selection process. This bias is further enhanced by the effects induced by using bootstrapping during the training stage [172]. An alternative unbiased method to standard random forests is the conditional inference forests methodology, that builds unbiased classification trees based on a conditional inference framework [169]. Unlike standard trees, unbiased ones employ Pearson $\chi 2$ test conditional p-values, corrected for multiplicity, which removes the bias towards multi-categorical predictors. Furthermore, the method employs *subsampling without replacing* during training, and a *conditional permutation importance* scheme during testing, which ensures an unbiased variable selection process [172].

3.2.11 Random forests applications

Genetic epidemiology is the medical field where random forest methods have been extensively applied, specifically in large-scale association studies, and patient outcome predictions from high-dimensional gene expression data [154]. Unfortunately, suicide research has not benefited from utilization of this powerful machine learning tool until very recently. To date, only three studies have used this methodology to help predicting fatal and non-fatal suicide attempts. One study focusing on predicting suicides after psychiatric hospitalization in US Army soldiers, used ensemble of regression trees as a preliminary step for selecting the most significant predictors in both bivariate associations and interactions [135]. In another study, random forests were used along with three other machine learning methods to select the optimal classifier for predicting suicides after outpatient mental health visits in Regular US military personnel [136]. Finally, random forests were employed as the sole machine learning method for predicting suicide attempts risk over time in civilians [137]. We are not aware of any Canadian suicide research studies employing this methodology.

4. Methods

4.1 Samples

Data sources were drawn from two cross-sectional population-based surveys: the 2012 Canadian Community Health Survey-Mental (CCHS-MH) [123], and the 2013 Canadian Forces Mental Health Survey (CFMHS) [122]. Both surveys adopted a sampling framework (stratified random sampling in CFMHS and multistage stratified cluster sampling in CCHS-MH) to ensure the representativeness of the samples in relation to their corresponding population. CCHS-MH included a representative sample of non-institutionalized individuals aged 15 years and older living in private dwellings in the 10 provinces (n=25113, response rate 69.8%). CFMHS included representative samples of serving CAF Regular Force personnel (n=6696, response rate 79.8%), and serving CAF Primary Reserve Force personnel who had deployed in support of the mission in Afghanistan (n=1469, response rate 78.7%), for a combined sample size of 8165 individuals. 14 modules of the survey content were commonly assessed in both surveys, however, some block were assessed in only one sample (Appendix).

4.2 Outcome measures

Suicidal ideation was assessed by asking all the respondents whether they had "seriously thought about committing suicide or taking [their] own life" and whether this experience had happened during the past 12 months. *Suicide plans* and *suicide attempts* were assessed by asking *lifetime ideators* whether they "made a plan for committing suicide" or "attempted suicide or tried to take [their] own life" and whether these experiences had happened during the past 12 months.

4.3 Covariates

Given that the conditional inference random forests methodology is capable of determining the optimal set of risk factors for suicidal behaviour on its own, without prior input from the researcher, we adopted a non-discriminative approach regarding which covariate qualifies as a candidate predictor. The only two inclusion criteria that each covariate must fulfill were: (1) to be assessed on both surveys and (2) to have a recall period of similar length or longer than that of the outcome. 225 covariates fulfilled both inclusion criteria and were included in the present analysis. These variables were derived from items pertaining to 14 modules commonly assessed in both surveys: General Health, Alcohol Use, Abuse, and Dependence, General Anxiety Disorder, Chronic Conditions, Medication Use, Perceived Need for Care with Help Needed sub-block, Work Stress, Social Provisions Scale, Depression, Suicide, Mental health Experiences, Mental Health Services, Adverse Childhood Experiences, and Socio-Demographics (including Dwelling and Household, Education, Income, and Labour force). Table A.1 describes how each variable was categorized.

4.3.1 Mental disorders diagnoses

Both surveys used the World Health Organization Composite International Diagnostic Interview (WHO-CIDI) to generate mental disorders diagnoses, based on Diagnostic and Statistical Manual of the American Psychiatric Association, fourth edition (DSM-IV) criteria [188]. WHO-CIDI is a lay-administered instrument with well-established consistency with clinical diagnostic instruments [189,190]. Only 3 mental disorders were assessed in both surveys: major depressive episode (MDE), generalized anxiety disorder (GAD), and alcohol use disorder (AUD). We distinguished between past-year MDE with and without significant distress. Additional self-reported mental disorders were assessed by asking the respondents if they had been diagnosed by a health care professional with a long-term health condition that lasted or was expected to last 6 months or longer: bipolar disorder, mania, dysthymia, phobia, obsessive-compulsive disorder (OCD), panic disorder, and PTSD.

4.3.2 General health

General health variables included perceived health, satisfaction with life in general, self-perceived health compared to one year ago, whether the respondent had trouble sleeping, perceived life stress, self-perceived work stress, and perceived sense of belonging to local community.

4.3.3 Alcohol use, abuse, and dependence

We used 20 items from the module on alcohol use, abuse, and dependence to derive 2 variables on past-year alcohol consumption (overall frequency, and frequency of 5 or more drinks on one occasion), 4 variables on lifetime alcohol abuse (whether alcohol abuse interfered with responsibilities, whether it was continued despite affecting others, whether the respondent could have been hurt, and how many police encounters did the respondent had due to drinking), 8 on lifetime alcohol dependence (having a strong urge to drink, needing more alcohol to get an effect, experiencing or drinking to avoid withdrawal symptoms, using more alcohol than intended, whether attempts were made for quitting, whether the respondent had periods spent mostly on drinking or recovering from the effects of drinking, whether drinking caused a significant reduction of important activities, and whether the respondent continued drinking despite causing serious physical or emotional problems), and 3 variables on alcohol-related healthcare services (lifetime consultation with healthcare professional, lifetime treatment, and lifetime hospitalization for alcohol problems).

4.3.4 General anxiety disorder

From the module on *general anxiety disorder*, we used 56 items to derive 25 variables describing topics causing excessive anxiety or worry (everything or nothing in particular, finances, success at school or work, social life, love life, relationships at school or work, relationship with family, physical appearance, own physical health, own mental health, alcohol or drugs, separation from loved ones, health or welfare of loved ones, social phobias, agoraphobia, specific phobias, obsessions, compulsions, crime or violence, economy, environment, moral decline of society, war or revolution, and other), 1 variable describing whether the excessive anxiety or worry lasted at least 6 months, 11 variables on anxiety-related symptoms (restlessness, fatigue, irritability, difficulty concentrating, tense or aching muscles, sleep problems, heart pounding, sweating, trembles or shakes, dry mouth, and sadness or depressed mood), 8 items describing the impact of anxiety symptoms (having difficulty controlling anxiety symptoms, whether clinically significant distress was present, intensity of emotional distress, lifetime degree of interference with daily activities, work, social life or personal relationships, intensity of past-year interference with home responsibilities, close relationships, and social life), 2 variables describing the number and type of anxiety episodes, including the most recent one, 3 past-year interference constructs (social or occupational interference, number of days totally unable to work or carry normal activities due to excessive anxiety, and overall interference), and 1 variable on anxiety-related healthcare services (lifetime consultation with healthcare professional, lifetime treatment, and lifetime hospitalization due to anxiety).

4.3.5 Chronic physical conditions

Chronic physical conditions were assessed by using a question about whether the subject had one of the following long-term conditions which are expected to last or have already lasted six months or more and that have been diagnosed by a health professional: asthma, arthritis, back problems (excluding fibromyalgia and arthritis), high blood pressure, migraine headaches, and diabetes.

4.3.6 Past-year medication use

Past-year medication use was assessed by asking the respondent whether during the past year prior to the interview received "any medication to help out with problems with your emotions, mental health or use of alcohol or drugs".

4.3.7 Perceived need for care

Overall *past-year perceived need for mental health care* was derived from 4 items indicating whether, during the past year preceding the interview, a perceived need for care was reported for problems with emotions, mental health or use of alcohol and drugs in the following domains: counselling, information, medication, and other help [191]. Using another 4 items from the same module on perceived need for care, we created a variable indicating whether the respondent received as much help as needed in one of the following forms: information, medication, and counselling or therapy.

4.3.8 Help needed

12 variables captured reasons for not seeking help after receiving partial help or no help at all in any of the aforementioned 3 domains: self-managed, unaware of help being available, too busy,

job interfered, help was not readily available, distrust in the system, help unaffordable, insurance not cover, fear of others' opinions, language problems, help was ongoing, and other reasons.

4.3.9 Work stress

From the work stress module, we used 1 item regarding job satisfaction, and 12 items describing self-reported perceptions about the following 5 dimensions of respondent's past-year work: security (job security was good), social support (whether supervisor or co-workers were helpful in getting the job done, and whether the respondent was exposed to hostility or conflict with others), skill discretion (job required learning new things, high level of skill, or repetitive tasks), physical exertion (job required lot of physical effort), and the extent of participation in decision making (job allowed for freedom to decide, and whether the respondent had own job input) [192,193].

4.3.10 Social support

Social support was assessed by asking the respondent about current relationships with friends, family members, co-workers, and community members. A shorter, 10-item version of the original Social Provisions Scale [194] was used to assess 5 main social provisions of social relationships: attachment (whether the respondent has close relationships or strong emotional bond), guidance (having someone to talk about important decisions, and whether the respondent has trustworthy persons for advice), reliable alliance (having people to depend on or people to count on emergency), social integration (having people who enjoy the same activities, and feeling as being part of a group with shared attitudes or beliefs), and reassurance of worth (having relationships the recognize competence, and having people who admire respondent's talents and abilities). An additional composite variable, measuring the overall score for the Social Provisions Scale, was created.

4.3.11 Depression

51 items from the depression module were used to derive 29 variables: 4 describing daily duration of a 2-week period of being sad, discouraged, or uninterested, the intensity of emotional distress associated with it, whether it was accompanied by severe emotional distress, and whether it included depressive-like symptoms, 2 variable describing the 2-week episode, 1 on bereavement, and 22 variables describing lifetime depressive-like symptoms experienced most of day during a period of 2 weeks or longer (being sad, empty or depressed, being discouraged or hopeless about the future, loosing interest in almost all things, feeling that nothing has fun, having significant weight loss, weight gain, or change in appetite, having insomnia or hypersomnia, fatigue, restlessness or retardation, diminished ability to think, trouble concentrating, being undecided, loosing self-confidence, feeling worthless, guilty, irritable, anxious, having attacks of fear or panic, being unable to cope with everyday responsibilities, being less talkative than usual, being often in tears, thinking of death, and experiencing passive suicidal ideation).

4.3.12 Suicide

From the suicide module, we generated 2 variables distinguishing between (1) past 12-month ideation with a history of lifetime ideation during a 2-week episode of depressive-like symptoms, such as being sad, depressed, or uninterested, and (2) past 12-month ideation with a history of lifetime ideation unrelated to such episodes. A similar pair of past 12-month suicide plans variables was created. We used these variables as covariates in subsequent models of suicide plans and suicide attempts, respectively.

4.3.13 Mental health experiences

Past-year mental health experiences were assessed by asking the respondent whether he/she worked in a mental health treatment program or with someone treated for a mental health problem, whether a family member, a friend, or himself/herself received mental health treatment during the past year preceding the interview, and whether he/she was discriminated due to own mental health, and subsequently impacted by stigma experiences.

4.3.14 Mental health services

Past-year hospitalization was assessed with 2 items indicating frequency and duration of hospitalizations for mental health problems. For each of the following 5 specialists, we created 4 variables describing past-year frequency of consultations, duration of average consultation, perceived level of help, and whether the respondent stopped seeing the specialist: psychiatrist, family doctor, psychologist, nurse, and social worker. For each of the following encounters, we created 2 variables describing past-year frequency of consultations, and perceived level of help: family, friend, co-worker or supervisor, and others. Past-year internet use for problems with emotions, mental health, alcohol or drugs was assessed with 5 items indicating whether the respondent used the internet to learn about symptoms, find out where to get help, discuss with others, get online therapy, or for other reasons. Experience with self-help groups was assessed by asking what type of self-help group the respondent went during the past year: emotional or mental health, alcohol or drug, and other. 2 items indicated whether the respondent felt prejudice or discrimination because of personal characteristics or own mental health problems during the past year. Finally, one item described whether the respondent needed help for emotions, mental health or use of alcohol or drugs during previous year but did not receive it.

4.3.15 Child abuse victimization

All respondents 18 years and older were asked to recall adverse childhood experiences that occurred before the age of 16 years, using items from the Childhood Experiences of Violence Questionnaire (CEVQ) [195]. Following CEVQ guidelines, 3 variables were constructed based on frequency of occurrence: exposure to domestic violence, physical abuse, and sexual abuse. An additional item, regarding lifetime contact with child protection organizations, was included.

4.3.16 Socio-demographic variables

Socio-demographic variables included age, sex, marital status, personal education, current living arrangement, racial background, sexual orientation, working status, personal income, and whether the respondent had difficulty meeting basic expenses with current household income.

4.4 Statistical analysis

4.4.1 Software

Data pre-processing was conducted by using SAS 9.3 and SAS Enterprise Guide 6.1 (SAS Institute Inc., Cary, NC, USA). All components of statistical modeling (data splitting, model tuning using resampling, variable selection, variable importance measures, and prediction) were performed in R version 6.2.2 [196] by using the *caret* package [197,198], and conditional inference random forests model with conditional permutation accuracy scheme from the *party* package [199]. *pROC* [200] and *ggplot2* packages were used to visualize the performance of classifiers. Samples of R-Code used for statistical modeling are provided in the Appendix. All computations were performed on standard desktop PC running 64-bit Windows 7 operating system, with Intel Xeon X5647 processor at 2.93GHz, and up to 96GB RAM.

4.4.2 Data pre-processing

Handing missing values

Unfortunately, the conditional permutation scheme cannot be applied to variables with missing values [168]. Complete case analysis and multiple imputation were not viable options to deal with this issue, since one would drastically reduce the number of instances available for training, while the other would inflate the current sample size several folds, with intractable consequences in terms of computational time. Therefore, we created a distinct category for all variables with missing values (don’t know/ refusal/ not stated).

Handling high dimensionality of the data

Handling missing data resulted in an increase in the dimensionality of the data. Applying a computationally demanding scheme such as conditional permutation to a dataset containing 225 variables with 455 categories (including missing values categories) presented a real challenge in terms of computational costs. We mitigated the additional burden associated with the high dimensionality of the data by collapsing multi-categorical variables before employing the algorithm. For each newly created dichotomous variable, the referent category is given in Table A.1. As a result, we presented the algorithm with 290 binary covariates for suicidal ideation, 294 for suicide attempts and 298 for suicide attempts (Table A.1). The difference in the number of covariates between models is due to the addition of suicidal ideation variables to suicide plans and attempts models, and suicide plans covariates to suicide attempts model. We used this strategy during hyper-parameter optimization, and model selection. However, most influential variables selected in the final models were used in their original format during the prediction stage.

4.4.3 Statistical modeling

For the present study, we relied on random forest methodology to select models of most influential predictors for suicidal ideation, suicide plans, and suicide attempts, and to evaluate their predictive performance on new subjects. We did not specify a model a priori but used all available covariates to let the algorithm learn on its own, through many iterations, which covariates are the most influential for the outcome.

Specifically, we applied conditional random forests methodology to mitigate the standard random forests selection bias towards highly-correlated variables [185]. Various settings of the random forests hyper-parameters were explored via a pre-defined tuning grid using empirical risk minimization [156]. We did not rely on *mtry* and *ntree* default values because they could not guarantee an optimal selection of classifiers [168]. Values corresponding to the model with the highest predictive accuracy were considered optimal and used during the model selection phase.

We addressed the class-imbalance problem by employing a *random sub-sampling within resampling* scheme [153], that is, before starting growing each tree, we *randomly resampled* the learning set and *under-sampled* the majority class to balance each training set resample. We preferred this method, in conjunction with the reduction in data dimensionality, to decrease the computational burden associated with using the conditional permutation scheme [201], and to ensure that the variation generated by subsampling is properly captured [153].

To mitigate the selection bias towards highly-correlated variables, we employed a *conditional permutation accuracy* scheme that allowed for a reliable estimation of variable importance scores, and a proper model selection [185]. Finally, to ensure that the predictions reflect the probability of suicidal ideation, suicide plans, and attempts observed in the real world, all validation or blind testing sets were kept unbalanced [153] (Figure 4.1).

4.4.4 Variable selection

For variable selection, we used a 2-step performance-based strategy which combines in a sequentially manner recursive elimination of least-important variables and stepwise forward addition of most influential ones [165,202]. First, we ranked the covariates according to their VIM scores, and eliminated the least-important variables, based on a threshold conditioned on their ranking. Any VIM score was measured in terms of the increase in misclassification rate after variable's relationship with the outcome was lost through random permutation [175]. Second, we used an ascendant strategy of adding most-important variables obtained previously, one at a time, according to their ranking, and creating nested models, whose predictive accuracies were subsequently compared (Figure 4.1). Our approach is an adaptation of the variable selection strategy developed by Genuer et al. [165,202]. However, it differs from it in several important ways. First, we relied on conditional inference forests instead on standard random forest methodology, which is known for its *variable selection bias* towards highly correlated predictors [185]. Second, we did not perform the thresholding for removing least-important variables on VIM mean, but on standard deviation of VIM, because we wanted to allow the algorithm to be able to shuffle the top ranked variables several times, before moving into selecting the most parsimonious model. Third, we used the step of eliminating least-important variables several times before moving to the variable selection stage, as we noticed that authors' assumption that variables ranking stabilizes after only one round of feature reduction did not hold with our data. Thus, we took advantage of current model variables ranking to adaptively update the candidate predictors distribution before next splitting [165]. Fourth, we relied on different settings during variable selection step, such as generating two times more iterations for computing OOB accuracy of nested models, and using the more stringent one standard error rule for selecting each of the less complex models [145].

4.4.5 Model performance

Model performance during hyper-parameter optimization and model selection was evaluated using accuracy. For prediction on external validation, we relied on AUC and metrics from the confusion matrix.

Figure 4.1. Main steps of conditional random forests algorithm

1. Define a tuning hyper-parameter grid;
2. Randomly split original data set into a training set and a blind testing set;
3. For each (*mtry*, *ntree*) set, and each tree:
    a. *Randomly* select a subset of *mtry* predictors;
    b. *Randomly* subsample *without replacement* from the training set obtained in step 2;
    c. *Randomly down-sample* the training subset at step 3b;
    d. Train on *balanced* sample at step 3c;
    e. Test on testing subset at step 3b and obtain predictive accuracy;
    f. Average *ntree* estimates to obtain mean predictive accuracy;
    g. Use conditional permutation scheme to obtain VIMs;
4. Run steps 3a to 3g 50 times to increase the stability of VIMs;
5. For each (*mtry*, *ntree*) set, obtain aggregate predictive accuracy (averaged over 50 iterations);

6. For each variable, obtain mean VIM (averaged over 50 iterations);
7. Select the model with the highest overall predictive accuracy;
8. Compare the model selected at step 7 to any *antecedent* model generated by the tuning grid:
    a. Evaluate the null hypothesis of no difference in predictive accuracy between models;
    b. Select the model with the smallest value of *ntree* and *mtry*, for which the null hypothesis cannot be rejected;
    c. Rank variables by sorting mean VIM in descending order;
    d. Remove *non-informative* predictors (VIM≤0);
9. Repeat steps 3 to 6;
10. Select the model, containing a subset of predictors, that has the highest aggregate predictive accuracy;
11. Compare the less complex model selected at step 10 to any *antecedent* model generated by the tuning grid:
    a. Evaluate the null hypothesis of no difference in predictive accuracy between models;
    b. Select the model with the smallest values of *ntree* and *mtry*, for which the null hypothesis cannot be rejected;
    c. Rank variables by sorting mean VIM in descending order;
    d. Remove *least-informative* predictors:
        i. Fit a conditional inference tree to estimate a threshold value of importance;
        ii. Retain only those predictors whose standard deviations of importance exceed the threshold value;
12. Repeat steps 9 to 11 until the predictive accuracy of the last less complex model drops below the accuracy of the best previous model minus its standard error;

13. Starting with the last complex model *within* one standard error of the previous best model, create nested models using variables sorted in descending order by mean VIM. Begin with the simplest model and do:
    a. Steps 3 to 5 (but skipping step 3g) to obtain its aggregate predictive accuracy;
    b. Discard the model if its predictive accuracy is below one standard error of the accuracy of the last complex model selected at step 13;
    c. Add to the model the next most important variable according to its ranking, and repeat 13a and 13b;
    d. Continue until no more variables can be added, or the model accuracy drops below one standard error of the model obtained at step 13;
14. Use 10-fold CV repeated 10 times to test the model obtained at step 13d:
    a. Retain the values of *mtry* and *ntree* from step 13d;
    b. Use variables selected at step 13d and train the algorithm on the training set obtained at step 2;
    c. Test the algorithm on the blind test set obtained at step 2;
    d. Use variables selected at step 13d and train the algorithm on the entire survey data set;
    e. Test the algorithm on the entire alternate survey data set;
    f. Generate and compare measures of performance against alternative measures from competing models.

The main steps of the procedure are presented in Figure 4.1. A detailed exposition of the computational steps we followed is given below:

1. Define a tuning hyper-parameter grid in which the default values of the number of variables to be chosen at each split *(mtry*=$\sqrt{p}$) and the number of trees in the forest (*ntree*=500) were sequentially increased by a *multiplicative factor* [153,164]. We employed this heuristic approach to allow the algorithm explore a higher number of possible predictors than the default values would allow. Recall that the chance for a predictor to enter the ensemble depends on the values of both hyper-parameters: the higher their values, the greater the chance for the predictor to be selected [168]. Setting the multiplicative factor to $2^{\mathbf{n}}$ allowed for a reasonable trade-off between performing a relatively thorough model selection while preserving computational efficiency, as conditional inference forests methodology is more computer-resource demanding than the standard approach [203]. Therefore, *mtry* and *ntree* are of the form $2^{\mathbf{n}}\cdot\sqrt{p}$, and $2^{\mathbf{n}}\cdot 500$, respectively, where **n** is a non-negative integer. Since the original datasets contain slightly less than 300 binary variables, and because the highest value for 2^n is the closest power of 2 not exceeding $\sqrt{p}$, it follows that $n\in\{0,1,2,3,4\}$, and $2^n\in\{1,2,4,8,16\}$ (Figure 4.2). Thus, *mtry* can take discrete values ranging from 1 in the most parsimonious model (p=2) to 272 in the full model. For *ntree*, we restrict the maximal value of **n** to 2 to prevent increased computational burden. Depending on the size of the feature space, the number of possible (*mtry*, *ntree*) tuning grid sets can range from 3 to 15. For example, with only 10 variables in the feature space, $\sqrt{p}$ =3, $2^n\in\{1,2\}$, *mtry*$\in\{3,6\}$, and the number

of possible (*mtry*, *ntree*) sets is 6: (3,500), (3,1000), (3,2000), (6,500), (6,1000), and (6,2000).

Figure 4.2. Tuning grid hyper-parameter setting values

n	Multiplicative factor 2^n	mtry	ntree
0	1	$\sqrt{p}$	500
1	2	$2\sqrt{p}$	1000
2	4	$4\sqrt{p}$	2000
3	8	$8\sqrt{p}$	-
4	16	$16\sqrt{p}$	-

2. Randomly split each original data set into (1) a training set to build the classifier for each outcome, and (2) a blind testing set to evaluate its predictive accuracy. We used equal sized sets for most of the outcomes. However, modeling suicide ideation in civilians on a training set half of the size of the original CCHS-MH dataset was not possible due to increasing computational burden associated with the subsequent conditional variable selection process. We used instead a training set of 25% of original data set. R software uses virtual memory for manipulating and holding data and there are limits on the amount of available memory [196]. A collection of data structures that exceeds this amount forces the process to stop. With 50% of the original CCHS-MH dataset used for training and subsequent variable selection, the operating system was unable to provide the necessary memory to build all the data structures. The execution of the process was terminated when the amount of available virtual memory was exceeded. Two main factors contributed to memory exhaustion: (1) the size of the training set, and (2) the use, over several iterations, of a conditional permutation scheme for variable importance measures. Also, we could not model suicide attempts in CAF using the hold-out method, due to the sparsity of the outcome in the CFMHS data set. Neither split worked. We

therefore relied on the full CFMHS sample and 10-fold CV repeated 10 times to model suicide attempts in CAF.

3. For each outcome, each (*mtry*, *ntree*) set, and each tree:
    a. *Randomly* select a subset of *mtry* predictors;
    b. *Randomly* subsample *without replacement* 62.3% of the observations from the training set obtained in step 2; Sampling without replacement is standard procedure in conditional random forests, ensuring that the selection bias observed with bootstrapping is not met [172];
    c. Subsample the training subset obtained in step 3b by *randomly down-sampling* the most prevalent class so that its frequency equals that of the least frequent class. Balancing avoids generating classifiers that do not accurately predict the minority class [153,156]. The main reason for using down-sampling instead of other methods (up-sampling, SMOTE) was to avoid the huge computational burden associated with using conditional permutation scheme at a subsequent step [201].
    d. Train the algorithm on balanced sample obtained in step 3c;
    e. Test the algorithm on 36.8% observations not selected in step 3b and obtain an estimate of predictive accuracy. Unlike training, the testing step is performed on an *unbalanced* set, ensuring that the procedure generates estimates reflecting the truthful performance of the algorithm [153];
    f. Average all *ntree* estimates to obtain mean estimate of predictive accuracy;
    g. Use conditional permutation scheme [185] to obtain variable importance measures (VIM) for all *p* variables. This method attenuates the correlation

between variables, and avoids the *selection bias* observed with unconditional permutation scheme, ensuring that the impact of each predictor on the outcome is correctly reflected [185];

4. Run steps 3a to 3g 50 times to increase the stability of VIM estimates [156,165,186]. Relying on a single algorithm iteration to report VIMs is generally discouraged [186], because VIMs and their rankings may vary with different random seed settings [185];
5. For each (*mtry*, *ntree*) set, obtain an aggregate estimate of predictive accuracy (averaged over previous 50 iterations);
6. For each variable, obtain a mean VIM estimate (averaged over previous 50 iterations);
7. Select the *best full model* (i.e. model containing all *p* variables that has the highest overall estimate of predictive accuracy);
8. Compare the difference in predictive accuracy between the *best full model* and all *antecedent full models* (i.e. full models for which the values of *mtry* and/or *ntree* are lower than the corresponding values of the *best full model*) generated by the tuning grid:
   a. evaluate the null hypothesis of no difference between the *best full model* and any *less performant antecedent full model* by means of corrected resampled t-tests [162,163]. The paired t-test is not suitable when averaging over many forests, because both learning and testing sets overlap across forests, leading to an underestimation of the variance and an increased probability of type I error [150,156]. For a significance level of α=0.05, and n-1 degrees of freedom, the null hypothesis can be rejected if $|t|>t_{49,\ 0.9725}$. Since we obtained each predictive accuracy by averaging over 50 forests, the threshold is $t_{49,\ 0.9725}$=2.0096. Two-

tailed p-values of less than 0.05 were considered to indicate statistical significance;

b. select the *optimal full model* (i.e. full model with the smallest values of *ntree* and *mtry*, for which the null hypothesis cannot be rejected). This step is essential in significantly reducing the computational time, because a classifier with less trees, but similar performance as the best one, can be subsequently utilized during model selection process;

c. Rank variables by sorting mean VIM estimates in descending order. The importance of a predictor is determined by the decrease in model's predictive accuracy after its values are permuted to destroy any relationship with the outcome. Larger differences in performance between models correspond to more important predictors and higher rankings;

d. Remove *non-informative* predictors, that is, predictors whose importance score is either negative or zero (VIM≤0) [Strobl 2009];

9. Repeat steps 3 to 6 once;

10. Select the *best reduced model* (i.e. model containing a subset of predictors that has the highest overall estimate of predictive accuracy);

11. Compare the difference in predictive accuracy between the *best reduced model* and all *antecedent reduced models* (i.e. reduced models for which the values of *mtry* and/or *ntree* are lower than the corresponding values of the *best reduced model*) generated by the tuning grid:

    a. Evaluate the null hypothesis of no difference between the *best reduced model* and any *less performant antecedent reduced model* by means of corrected resampled t-tests. For a significance level of $\alpha=0.05$, and n-1 degrees of freedom, the null hypothesis can be rejected if $|t|>t_{49,\ 0.9725}$. Since we obtained each predictive accuracy by averaging over 50 forests, the threshold is $t_{49,\ 0.9725}=2.0096$. Two-tailed p-values of less than 0.05 were considered to indicate statistical significance;

b. Select the *optimal reduced model* (i.e. reduced model with the smallest values of *ntree* and *mtry,* for which the null hypothesis cannot be rejected);

c. Rank variables by sorting mean VIM estimates in descending order;

d. Reduce feature space by removing *least-informative* predictors:

   i. Using ordered sequence of standard deviations of VIM, fit a conditional inference tree to estimate a threshold value of importance by predicting the standard deviation of importance of least-informative predictors (i.e. variables with less variability) [165,202];

   ii. Retain only those predictors whose standard deviations of importance exceed the threshold value;

12. Repeat steps 9 to 11 until the predictive accuracy of the last *reduced model* drops below the accuracy of the best previous reduced model minus its standard error [145]. The earliest baseline model for measuring predictive performance was the classifier constructed on non-redundant variables as we expected a decrease in accuracy larger than one standard error after the removal of non-informative variables [164,166].

13. Starting with the *last best reduced model* (i.e. last reduced model *within* one standard error of the previous best reduced model), create *nested models* using variables sorted in

descending order, according to their mean VIM. Begin with the simplest nested model (containing the most important 2 variables) and do:

a. Steps 3 to 5 (but skipping step 3g) to obtain its aggregate predictive accuracy;
b. Discard the model if its predictive accuracy is below one standard error of the accuracy of the *last best reduced model*;
c. Add to the model the next most important variable according to its ranking, and repeat 13a and 13b;
d. Continue until no more variables can be added, or the model accuracy is within one standard error of *last best reduced model*;

14. Use 10-fold CV repeated 10 times to perform within-sample and between-sample external validation of the *final model* (i.e. model obtained at step 12d):
    a. Retain the values of *mtry* and *ntree* from step 13d;
    b. Use variables selected at step 13d, reverted to their original categories, and train the algorithm on the training set at step 2;
    c. Test the algorithm on the blind test set at step 2 (i.e. within-sample external validation);
    d. Test the algorithm on the entire data set of the other survey (i.e. between-sample external validation);
    e. Generate measures of performance, and compare them against alternative measures from competing classifiers, by means of corrected resampled t-tests [162,163].

5. Results

We used conditional inference random forests algorithm to build models of suicidal behavior in Canadian general population and CAF. Optimal values of *mtry* and *ntree* hyper-parameters were chosen by means of a pre-specified tuning grid of sufficient size that would preserve the computational efficiency of the procedure. All training datasets were imbalanced, due to the rarity of the outcomes. To solve the imbalance problem, we employed a random sub-sampling during resampling scheme, in which each training dataset was balanced by down-sampling the majority class. However, the performance of each model was evaluated on unbalanced testing sets, to reflect the true prevalence of the outcome in the population. We employed a conditional permutation accuracy scheme to obtain an optimal selection of the most influential cross-sectional predictors on the suicidal behaviour outcomes. This method attenuates the correlation between variables, ensuring that the impact of each predictor on the outcome is correctly reflected. Finally, to ensure that the predictions reflect the probability of suicidal behaviour outcomes observed in the real world, we kept all validation or blind testing sets unbalanced.

5.1 Hyper-parameter optimization

During hyper-parameter optimization step, we trained most of the conditional inference random forests classifiers on datasets half the size of their corresponding original samples. Hold-out sampling method was used to obtain the predictive performance of each model, except for the one predicting suicide attempts in CAF. Due to the rarity of the outcome in this population, we relied on 10-fold cross validation repeated 10 times to obtain the performance of the model. By using the entire sample, this method is suitable for situations in which the split of the sample size cannot ensure a proper training of the classifier.

5.1.1 Hyper-parameter optimization of conditional inference random forests models of suicidal behaviour in Canadian general population

Suicidal ideation

The performance of the full model (including all 290 predictors) of suicidal ideation in Canadian general population is shown in Tables 7.1 and 7.2 and Figure 7.1. The predictive accuracy averaged over 50 iterations ranged from 74.53% with default values for *ntree* and *mtry* to 77.82% in models built on default value of *ntree* and a value for *mtry* 8 times larger than its default. Increasing the default value of *mtry* by a factor up to 8, while keeping *ntree* fixed, resulted in an absolute increase in predictive accuracy. However, further increase of *mtry* resulted in a decrease in performance. On the other hand, increasing *ntree*, while keeping *mtry* fixed, did not translate into an improvement in predictive accuracy.

Suicide plans

The performance of the full model (including all 294 predictors) of suicide plans in Canadian general population is shown in Tables 7.1 and 7.2 and Figure 7.2. The predictive accuracy averaged over 50 iterations ranged from 80.73% with default values for *ntree* and *mtry* to 94.25% in models built on default value of *ntree* and a value for *mtry* 16 times larger than its default. Increasing the default value of *mtry* by a factor up to 16, while keeping *ntree* fixed, resulted in an absolute increase in predictive accuracy. On the other hand, increasing *ntree*, while keeping *mtry* fixed, did not translate into an improvement in performance.

Suicide attempts

The performance of the full model (including all 298 predictors) of suicide attempts in Canadian general population is shown in Tables 7.1 and 7.2 and Figure 7.3. The predictive accuracy averaged over 50 iterations ranged from 80.39% with default values for *ntree* and *mtry* to 86.61% in models built on the maximal value for *ntree* and a value for *mtry* 8 times larger than its default. Increasing the default value of *mtry* by a factor up to 8, while keeping *ntree* fixed, resulted in an absolute increase in predictive accuracy. However, further increase of *mtry* resulted in a decrease in performance. On the other hand, increasing *ntree,* while keeping *mtry* fixed, did not translate into an improvement in predictive accuracy.

5.1.2 Hyper-parameter optimization of conditional inference random forests models of suicidal behaviour in Canadian Armed Forces

Suicidal ideation

The performance of the full model (including all 289 predictors) of suicidal ideation in CAF is shown in Tables 7.1 and 7.2 and Figure 7.4. The predictive accuracy averaged over 50 iterations ranged from 73.12%, with default value for *mtry* and a value for *ntree* double its default value, to 77.74% in random forests with values of *mtry* and *ntree* 8 several times larger than their default values. Increasing the default value of *mtry* by a factor up to 8, while keeping *ntree* fixed, resulted in an absolute increase in predictive accuracy. However, further increase of *mtry* resulted in a decrease in performance. On the other hand, increasing *ntree*, while keeping *mtry* fixed, did not translate into an improvement in predictive accuracy.

Suicide plans

The performance of the full model (including all 293 predictors) of suicide plans in Canadian Armed Forces is shown in Tables 7.1 and 7.2 and Figure 7.5. The predictive accuracy averaged over 50 iterations ranged from 79.04%, with default value for *mtry* and a value for *ntree* 4 times larger than its default value, to 86.46% in random forests with maximal grid pre-specified values of *mtry* and *ntree*. Increasing the default value of *mtry* while keeping *ntree* fixed, resulted in an absolute increase in performance. On the other hand, increasing *ntree*, while keeping *mtry* fixed, did not translate into an improvement in predictive accuracy.

Suicide attempts

The performance of the full model (including all 297 predictors) of suicide attempts in CAF is shown in Tables 7.1 and 7.2 and Figure 7.6. The predictive accuracy averaged over 50 iterations ranged from 76.26%, with default value for *mtry* and a value for *ntree* 2 times larger than its default value, to 80.76% in random forests with maximal values of *mtry* and *ntree* as specified by the pre-defined grid. Increasing the default value of *mtry* while keeping *ntree* fixed, resulted in an absolute increase in performance. On the other hand, increasing *ntree*, while keeping *mtry* fixed, did not translate into an improvement in predictive accuracy.

5.1.3 Overall results

Regardless of outcome and population setting, increasing the default value of *mtry* resulted in an absolute increase of the predictive accuracy of the full model. On the other hand, any increase of *ntree* did not translate into an improvement in predictive accuracy. Therefore, we decided to reduce the size of the tuning grid by relying on the default value for *ntree* only, while the setting values for *mtry* remained unchanged. This simplified approach helped avoiding computational burden associated with the use of large random forests of 1000 or 2000 trees each during the variable selection step.

5.2 Model reduction

During the model reduction step, we kept the number of trees in each forest fixed at the default value of 500, as larger forests did not exhibit better predictive accuracies during the hyper-parameter optimization (Tables 7.1 and 7.2). We trained most of the full conditional inference random forests models on datasets half the size of their corresponding original samples. However, we trained all reduced models of suicidal ideation in Canadian general population on only 25% of the original sample size, because the operating system was unable to provide the necessary memory to build the data structure generated by the conditional permutation scheme during the subsequent variable selection step. Also, we relied on the entire military sample to build models of suicide attempts in CAF, because splitting the sample size did not ensure a proper training of the classifier, due to the rarity of the outcome in this sample. Hold-out sampling method was used to obtain the predictive performance of each model, except for the one predicting suicide attempts in CAF, in which we relied on 10-fold cross validation repeated 10 times.

5.2.1 Predictive performance of reduced conditional inference random forest models of suicidal behaviour in Canadian general population

Suicidal ideation

The predictive performance of reduced conditional inference random forests models of suicidal ideation in Canadian general population is shown in Table 7.3 and Figure 7.7. The predictive accuracy ranged between 78.44%, in an intermediary reduced model with 115 predictors and a value for *mtry* 8 times higher than its default, to 80.32% in the best reduced model with 14 variables and a *mtry* value 4 times higher than its default. Removing non-informative predictors resulted in a decrease in performance, which was greater than 1 standard error of the predictive accuracy of the full model. However, further reduction in the feature space resulted in intermediary models whose accuracies were higher than or within 1 standard error of the predictive accuracy of their corresponding preceding reduced models. When compared to the first reduced model, from which all non-informative predictors were removed, the predictive accuracy of the best reduced model increased by 2.31%, while its feature space was reduced by 91.08%.

Suicide plans

The predictive performance of reduced conditional inference random forests models of suicide plans in Canadian general population is shown in Table 7.3 and Figure 7.8. The predictive accuracy ranged between 94.28%, in an intermediary reduced model with 53 predictors and a value for *mtry* 4 times higher than its default, to 94.64% in the best reduced model with 34 variables and a value for *mtry* 4 times higher than its default. Removing non-informative predictors as well as further reduction in the feature space resulted in intermediary models whose accuracies were higher than or within 1 standard error of the predictive accuracy of their corresponding preceding reduced models. When compared to the first reduced model, from which all non-informative predictors were removed, the predictive accuracy of the best reduced model increased by 0.35%, while its feature space was reduced by 72.80%.

Suicide attempts

The predictive performance of reduced conditional inference random forests models of suicide attempts in Canadian general population is shown in Table 7.3 and Figure 7.9. Removing non-informative predictors resulted in a predictive accuracy of 86.14%. This was an absolute increase in performance over the full model that was not exceeded by subsequent reduction in feature space. In other words, the first reduced model, with 99 predictors and a value for *mtry* equal to 4 times its default value, was also the best reduced model.

5.2.2 Predictive performance of reduced conditional inference random forest models of suicidal behaviour in Canadian Armed Forces

Suicidal ideation

The predictive performance of reduced conditional inference random forests models of suicidal ideation in CAF is shown in Table 7.4 and Figure 7.10. The predictive accuracy ranged between 77.40%, in the first reduced model with 141 predictors and a value for *mtry* 8 times higher than its default, to 79.02% in the best reduced model with 5 variables and a *mtry* value 2 times higher than its default. Removing non-informative predictors resulted in a decrease in performance, which was greater than 1 standard error of the predictive accuracy of the full model. However, further reduction in the feature space resulted in subsequent models whose accuracies were higher than or within 1 standard error of the predictive accuracy of their corresponding preceding reduced models. When compared to the first reduced model, from which all non-informative predictors were removed, the predictive accuracy of the best reduced model increased by 2.09%, while its feature space was reduced by 96.45%.

Suicide plans

The predictive performance of reduced conditional inference random forests models of suicide plans in CAF is shown in Table 7.4 and Figure 7.11. Removing non-informative predictors resulted in a decrease in performance, which was greater than 1 standard error of the predictive accuracy of the full model. However, further reduction in feature space did not result into an improvement of performance over the first reduced model. In other words, the first reduced model, with 92 predictors, a value for *mtry* equal to 8 times its default value, and a predictive accuracy of 85.52% was also the best reduced model.

Suicide attempts

The predictive performance of conditional inference random forests reduced models of suicide attempts in CAF is shown in Table 7.4 and Figure 7.12. The predictive accuracy ranged between 80.73%, in an intermediary reduced model with 105 predictors and a value for *mtry* 8 times higher than its default, to 83.78% in the best reduced model, with 4 variables and default *mtry* value. Removing non-informative predictors as well as further reduction in the feature space resulted in subsequent models whose accuracies were higher than or within 1 standard error of the predictive accuracy of their corresponding preceding reduced models. When compared to the first reduced model, from which all non-informative predictors were removed, the predictive accuracy of the best reduced model increased by 3.05%, while its feature space was reduced by 96.19%.

5.2.3 Overall results

Regardless of outcome and population setting, model reduction through sequential reduction of the feature space resulted in an overall increase in predictive accuracy. In most cases, removing non-informative predictors resulted in a slight decrease in performance, which was greater than 1 standard error of the predictive accuracy of the full model. However, further reduction in the feature space resulted in intermediary models whose accuracies were higher than or within 1 standard error of the predictive accuracy of their corresponding preceding reduced models. Moreover, all best reduced models, which were based on subsets of predictors representing much smaller feature spaces than the original ones, performed better than their corresponding best full models.

5.3 Feature space reduction and variable selection

For every outcome, attempts have been made to employ the best full model during the variable selection step. However, this approach led to an exhaustion of virtual memory. The operating system was unable to provide the necessary memory to build the data structure associated with employing conditional permutation scheme for variable selection. R software holds all data structures in virtual memory, and there are limits on the amount that can be allocated to them. Relying on the best full model created data structures that exceeded the available virtual memory.

To avoid increasing memory requirements, corrected resampled t-tests were employed to choose an optimal antecedent full model, defined as the least performant antecedent full model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best full model cannot be rejected. The optimal antecedent full classifier was subsequently used to generate variable importance scores and remove non-informative predictors.

To improve computational efficiency during each subsequent feature reduction step, corrected resampled t-test were employed to choose an optimal antecedent reduced model, defined as the least performant antecedent reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected. The optimal antecedent reduced model was subsequently used to generate variable importance scores and remove least-informative predictors.

The importance of a predictor was determined by the decrease in model's mean predictive accuracy after its relationship with the outcome was lost through random permutation. Larger differences in permutation accuracy importance scores corresponded to more important variables and higher rankings.

5.3.1 Feature space reductions and variable selection in conditional inference random forest models of suicidal behaviour in Canadian general population

Suicidal ideation

Table 7.5 and Figure 7.13 contrasts the permutation accuracy importance scores of the full model of suicidal ideation in Canadian general population to those from subsequent reduced models. A total of 10 reduced models were generated, resulting in a 95.17% decrease of the original feature space. 143 out of 290 predictors (49.31%) were non-informative and were removed from subsequent models. 7 predictors from 5 blocks (Mental Health Services (1), Medication (1), Mental Health Experiences (1), Depression (1), and GAD (3)) ranked among the top 10 most important predictors in all models, albeit in different order of ranking. Top 2 predictors kept their ranking throughout all models. The best reduced model included 14 predictors from 7 blocks: Mental Health Services (2), Medication (1), Depression (2), Mental Health Experiences (1), GAD (5), Alcohol Use Disorders (1), and Social Provision Scale (2).

Suicide plans

Table 7.6 and Figure 7.14 contrasts the permutation accuracy importance scores of the full model of suicide plans in Canadian general population to those from subsequent reduced models. A total of 3 reduced models were generated, resulting in an 88.44% decrease of the original feature space. 169 out of 294 predictors (57.48%) were non-informative and were removed from subsequent models. 9 predictors from 4 blocks (Suicide (2), Depression (5), Mental Health Services (1), and Medication (1)) ranked among the top 10 most important predictors in all models, albeit in different order of ranking. Top 4 predictors kept their ranking throughout all models. The best reduced model included 34 predictors from 8 blocks: Suicide (3), Depression (15), Mental Health Experiences (1), Mental Health Services (4), Medication (1), GAD (6), Adverse Childhood Experiences (2), Perceived Need for Care (1), and Chronic Conditions (1).

Suicide attempts

Table 7.7 and Figure 7.15 contrasts the permutation accuracy importance scores of the full model of suicide attempts in Canadian general population to the one from the subsequent reduced model. 199 out of 298 predictors (66.78%) were non-informative and were removed from the subsequent model. 9 predictors from 4 blocks (Mental Health Experiences (1), Medication (1), Suicide (3), and Depression (4)) ranked among the top 10 most important predictors in all models, albeit in different order of ranking. Top 7 predictors kept their ranking in both models. The best reduced model included 99 predictors from 13 blocks: Mental Health Experiences (6), Medication (1), Depression (30), Suicide (5), Perceived Need for Care (3), GAD (25), Mental Health Services (7), Socio-Demographics (7), Chronic Conditions (5), Alcohol Use disorders (4), Adverse Childhood Experiences (2), Work Stress (2), Social Provision Scale (1), and General Health (1).

5.3.2 Feature space reduction and variable selection in conditional inference random forest models of suicidal behaviour in Canadian Armed Forces

Suicidal ideation

Table 7.8 and Figure 7.16 contrasts the permutation accuracy importance scores of the full model of suicidal ideation in CAF to those from subsequent reduced models. A total of 9 reduced models were generated, resulting in a 98.27% decrease of the original feature space. 148 out of 289 predictors (51.21%) were non-informative and were removed from subsequent models. 4 predictors from 4 blocks (Mental Health Experiences (1), GAD (1), Mental Health Services (1), and Depression (1)) ranked among the top 10 most important predictors in all models, albeit in different order of ranking. The best reduced model included 5 predictors from 5 blocks: Mental Health Experiences (1), GAD (1), Mental Health Services (1), Depression (1), and Alcohol Use Disorders (1).

Suicide plans

Table 7.9 and Figure 7.17 contrasts the permutation accuracy importance scores of the full model of suicide plans in CAF to the one from the subsequent reduced model. 201 out of 294 predictors (68.60%) were non-informative and were removed from the subsequent model. Both models shared the same top 10 predictors, albeit in different order of ranking. These predictors were part of the following 5 blocks: Suicide (2), Depression (5), GAD (1), Mental Health Experiences (1), and Medication (1). Top 4 predictors kept their ranking in both models. The best reduced model included 92 predictors from 8 blocks: Suicide (3), Depression (29), GAD (22), Mental Health

Experiences (4), Medication (1), Perceived Need for Care (3), Mental Health Services (9), Chronic Conditions (5), Work Stress (5), Alcohol Use Disorders (6), Socio-demographics (3), Social Provisions Scale (2), and Adverse Childhood Experiences (1).

Suicide attempts

Table 7.10 and Figure 7.18 contrasts the permutation accuracy importance scores of the full model of suicide attempts in CAF to those from subsequent reduced models. A total of 8 reduced models were generated, resulting in a 98.65% decrease of the original feature space. 192 out of 297 predictors (64.65%) were non-informative and were removed from subsequent models. The best reduced model included 4 predictors from 4 blocks (GAD (1), Mental Health Experiences (1), Suicide (1), and Medication (1)). These variables also ranked among the top 10 most important predictors in all models, albeit in different order of ranking.

5.3.3 Overall results

At least one out of every two predictors was non-informative, regardless of the outcome and population setting. The proportion of non-informative predictors ranged between 49.31% in the first reduced model of suicide ideation in Canadian general population to 68.60% in the first reduced model of suicide plans in CAF. The subsequent elimination of least-informative predictors yielded best reduced models with sets of predictors ranging from 1.35% (in the best reduced model of suicide attempts in CAF) to 31.40% (in the best reduced model of suicide plans in CAF) of the corresponding initial feature spaces.

Among all predictors derived from the best reduced models of suicidal behaviour in Canadian general population, 9 were common to all 3 outcomes (Table 7.18) and had their source the following 5 modules: Mental Health Services (2), Medication Use (1), Depression (2), Mental

Health Experiences (1), and GAD (3). Among all predictors derived from the best reduced models of suicidal behaviour in CAF, 2 were common to all 3 outcomes (Table 7.18) and had their source in the following 2 modules: Mental Health Experiences (1), and GAD (1). Among all predictors of suicidal ideation, 3 from 3 modules (Mental Health Services, Mental Health Experiences, and GAD) were common to both population settings (Table 7.19). Among all predictors of suicide plans, 28 from 8 modules (Mental Health Services (2), Medication Use (1), Depression (14), Mental Health Experiences (1), GAD (6), Suicide (2), Adverse Childhood Experiences (1), and Perceived Need for Care (1)) were common to both population settings (Table 7.19). Among all predictors of suicide attempts, 4 from 4 modules (Medication Use (1), Mental Health Experiences (1), GAD (1), and Suicide (1)) were common to both population settings (Table 7.19).

5.4 Final model selection

During the final model selection step, we relied on a stepwise forward addition strategy to create nested models of suicidal behaviour in Canadian general population and CAF (Tables 7.13 to 7.16, and Figures 7.19 to 7.24). Briefly, we added the most influential predictors derived from the best reduced models and sorted by their permutation accuracy scores, one at a time, to an initial nested model of 2 most influential predictors. We determined the importance of a predictor based on the mean decrease in model's predictive accuracy after predictor's values were permutated to destroy its relationship with the outcome: the higher the decrease, the more influential the predictor.

5.4.1 Predictive performance of nested conditional inference random forest models of suicidal behaviour in Canadian general population

Suicidal ideation

The predictive performance of conditional inference random forests nested models of suicidal ideation in Canadian general population is shown in Table 7.11 and Figure 7.19. The predictive accuracy of any nested model with at most 8 variables was less than 1 standard error of the predictive accuracy of the best reduced model. However, the best nested model, with 9 variables (representing a reduction of 96.90% of the original feature space) and a value of *mtry* twice its default, yielded a predictive accuracy of 80.55%, which was higher than the one of the best reduced model, with 14 predictors and a value of *mtry* 4 times its default (80.32%). This model had the highest predictive accuracy among all models of suicidal ideation in Canadian general population.

Suicide plans

The predictive performance of conditional inference random forests nested models of suicide plans in Canadian general population is shown in Table 7.12 and Figure 7.20. The predictive accuracy of any nested model with at most 5 variables was less than 1 standard error of the predictive accuracy of the best reduced model. However, the best nested model, with 6 variables (representing a reduction of 97.96% of the original feature space) and a value of *mtry* twice its default, yielded a predictive accuracy of 94.80%, which was higher than the one of the best reduced model, with 34 predictors and a value of *mtry* 4 times its default (94.64%). This model had the highest predictive accuracy among all models of suicide plans in Canadian general population.

Suicide attempts

The predictive performance of conditional inference random forests nested models of suicide attempts in Canadian general population is shown in Table 7.13 and Figure 7.21. The predictive accuracy of any nested model with at most 5 variables was less than 1 standard error of the predictive accuracy of the best reduced model. However, the best nested model, with 6 variables (representing a reduction of 97.99% of the original feature space) and a value of *mtry* equal to its default, yielded a predictive accuracy of 86.46%, which was higher than the one of the best reduced model, with 99 predictors and a value of *mtry* 4 times its default (85.23%). This model had the highest predictive accuracy among all models of suicide attempts in Canadian general population.

5.4.2 Predictive performance of nested conditional inference random forest models of suicidal behaviour in Canadian Armed Forces

Suicidal ideation

The predictive performance of conditional inference random forests nested models of suicidal ideation in CAF is shown in Table 7.14 and Figure 7.22. The predictive accuracy of any nested model was less than 1 standard error of the predictive accuracy of the best reduced model. This model, with 5 variables (representing a reduction of 98.27% of the original feature space) and a value of *mtry* twice its default, yielded a predictive accuracy of 79.02%, which was highest among all models of suicidal ideation in CAF.

Suicide plans

The predictive performance of the only conditional inference random forests nested model of suicide plans in CAF is shown in Table 7.15 and Figure 7.23. This model, with 2 variables (representing a reduction of 99.32% of the original feature space) and a default value for *mtry*, yielded a predictive accuracy of 90.67%, which was the highest among all models of suicide plans in CAF, including the best reduced model, with 92 predictors and a value of *mtry* 8 times its default.

Suicide attempts

The predictive performance of conditional inference random forests nested models of suicide attempts in CAF is shown in Table 7.16 and Figure 7.24. The predictive accuracy of any nested model was less than 1 standard error of the predictive accuracy of the best reduced model. This model, with 4 variables (representing a reduction of 98.65% of the original feature space) and a value of *mtry* twice its default, yielded a predictive accuracy of 83.78%, which was highest among all models of suicidal ideation in CAF.

5.4.3 Overall results

Regardless of outcome, final nested models of suicidal behaviour in Canadian general population yielded similar or slightly higher predictive accuracies than their corresponding best reduced models, while relying on less predictors. Final models of suicidal behaviour in CAF had a more nuanced behaviour. For example, the initial nested model of suicide plans, with only 2 predictors, had a higher accuracy than the best reduced model of suicide plans, which relied on 92 predictors. However, in the other two instances all nested models had predictive performances which were lower than 1 standard error of the accuracy of their corresponding reduced models. In other words, best reduced models of suicidal ideation and suicide attempts outperformed any nested model.

5.5 Selection of final risk algorithms

The most influential cross-sectional risk algorithms of suicidal behaviour in Canadian general population and CAF are shown in Tables 7.20 and 7.21, respectively. Selection of a predictor in the final model was determined by its permutation accuracy score in the best reduced model and the performance of the nested model once the predictor has been added to it. Thus, any new added predictor was kept only if the predictive accuracy improved over the previous model.

5.5.1 Cross-sectional risk algorithms of suicidal behaviour in Canadian general population

Suicide ideation

The most influential cross-sectional risk algorithms of suicidal ideation in Canadian general population were constituted of 9 constructs from 5 blocks: Mental Health Services (Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it during the past 12 months), Medication Use (Received medication for emotions, mental health or use of alcohol or drugs during the past 12 months), Depression (Thought it would be better if dead; Thought a lot about own death or death in general), GAD (Often nervous or worried that could not think about something else; Worried about physical appearance; Worried about own mental health), Mental Health Experiences (Received treatment for emotional or mental health problem during past 12 months), and Alcohol Use Disorders (Unable to cut down alcohol use despite making several attempts to quit).

Suicide plans

The most influential cross-sectional risk algorithms of suicide plans in Canadian general population were constituted of 6 constructs from 3 blocks: Suicide (Past 12-month suicidal ideation and history of lifetime suicidal ideation during and outside worst depressive episode), Depression (Thought it would be better if dead; Thought a lot about own death or death in general; Felt so sad that nothing could cheer up), and Mental Health Services (Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it during the past 12 months).

Suicide attempts

The most influential cross-sectional risk algorithms of suicide attempts in Canadian general population were constituted of 6 constructs from 4 blocks: Mental Health Experiences (Received treatment for emotional or mental health problem during past 12 months), Medication Use (Received medication for emotions, mental health or use of alcohol or drugs during the past 12 months), Suicide (Past 12-month suicidal ideation and history of lifetime suicidal ideation during and outside worst depressive episode), and Depression (Thought it would be better if dead; Major depressive episode with significant interference on daily activities and responsibilities).

5.5.2 Cross-sectional risk algorithms of suicidal behaviour in Canadian Armed Forces

Suicidal ideation

The most influential cross-sectional risk algorithms of suicidal ideation in CAF were constituted of 5 constructs from 5 blocks: Mental Health Experiences (Received treatment for emotional or mental health problem during past 12 months), GAD (Worried about own mental health), Mental Health Services (Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it during the past 12 months), Alcohol Use Disorders (Unable to cut down alcohol use despite making several attempts to quit), and Depression (Major depressive episode with significant interference on daily activities and responsibilities).

Suicide plans

The most influential cross-sectional risk algorithms of suicide plans in CAF were constituted of 2 constructs from 2 blocks: Suicide (Past 12-month suicidal ideation and history of lifetime suicidal ideation outside the worst depressive episode), and Depression (Major depressive episode with significant interference on daily activities and responsibilities).

Suicide attempts

The most influential cross-sectional risk algorithms of suicide attempts in CAF were constituted of 4 constructs from 4 blocks: GAD (Worried about own mental health), Mental Health Experiences (Received treatment for emotional or mental health problem during past 12 months), Suicide (Past 12-month suicidal ideation and history of lifetime suicidal ideation during worst depressive episode), and Medication Use (Received medication for emotions, mental health or use of alcohol or drugs during the past 12 months).

5.5.3 Overall results

Regardless of population setting, the size and the degree of heterogeneity of the risk algorithms of suicide ideation were the largest among the three suicidal behaviour outcomes. More predictors from more domains were used to predict suicidal ideation than suicide plans or attempts. Lifetime passive suicide ideation (Thought it would be better if dead) was the only common predictor of all 3 risk algorithms of suicidal behaviour in Canadian general population. No such construct was identified for the military. However, other predictors were commonly found in several risk algorithms, regardless of the population setting. For example, feeling that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it during the past 12 months, past 12-month mental health treatment, and being worried about own mental health were common to both risk algorithms of suicide ideation in Canadian general population and CAF. Likewise, past 12-month mental health treatment and past 12-month medication for mental health issues were common to both risk algorithms of suicide attempts. Also, past 12-months suicide ideation was included in the risk algorithms of suicide plans, and suicide attempts, regardless of population setting.

5.6 External validation of final models

We performed the validation of the final conditional inference random forests models on equally sized unbalanced samples from the same population domains (within-sample external validation) as the ones on which we trained the classifiers as well on the other population setting (between-sample external validation). In other words, for each suicidal behaviour outcome, we used predictors derived from the best nested models to train the classifier on either (1) half of a sample of a certain population domain and test on the other half or to train on (2) the entire sample of a certain population domain and tested on the other entire population sample. While trained on balanced samples, all 6 final models were tested on unbalanced samples, to reflect the real prevalence of the suicidal behaviour outcomes. The test set performance measures are shown in Table 7.17 and in Figures 7.25 to 7.36.

5.6.1 Suicidal ideation

A conditional inference random forests model of suicidal ideation, containing 9 variables, yielded an accuracy, sensitivity, specificity, and AUC of 81.98%, 79.54%, 82.09%, and respectively 0.8812 when trained on the entire Canadian general population sample (n=25034), and tested on the entire CAF sample (n=8152). Training the same model on half of the Canadian general population sample (n=12519) and tested on the other half yielded higher accuracy (88.26%), and specificity (88.26%), but lower sensitivity (64.10%), and AUC (0.8505). On the other hand, a conditional inference random forests model of suicidal ideation, containing 5 variables, yielded an accuracy, sensitivity, specificity, and AUC of 92.95%, 51.45%, 94.55%, and respectively 0.8391, when trained on the entire CAF sample (n=8152), and tested on the entire CAF sample (n=25034). Training the same model on half of the CAF sample (n=4077) and tested on the other half yielded lower accuracy (88.91%), and specificity (89.98%), but higher sensitivity (64.94%), and AUC (0.8728).

5.6.2 Suicide plans

A conditional inference random forests model of suicide plans, containing 6 variables, yielded an accuracy, sensitivity, specificity, and AUC of 95.33%, 58.27%, 95.91%, and respectively 0.9520 when trained on the entire Canadian general population sample (n=25034), and tested on the entire CAF sample (n=8152). Training the same model on half of the Canadian general population sample (n=12517) and tested on the other half yielded higher estimates for accuracy (97.79%), sensitivity (68.97%), specificity (98.12%), and AUC (0.9793). On the other hand, a conditional inference random forests model of suicide plans, containing 2 variables, yielded an accuracy, sensitivity, specificity, and AUC of 95.40%, 58.24%, 95.83%, and respectively 0.8580, when trained on the entire CAF sample (n=8152), and tested on the entire CAF sample (n=25034). Training the same model on half of the CAF sample (n=4076) and tested on the other half yielded lower accuracy (93.08%), and specificity (93.52%), but higher sensitivity (67.16%), and AUC (0.9072).

5.6.3 Suicide attempts

A conditional inference random forests model of suicide attempts, containing 6 variables, yielded an accuracy, sensitivity, specificity, and AUC of 94.30%, 82.61%, 94.33%, and respectively 0.9786 when trained on the entire Canadian general population sample (n=25036), and tested on the entire CAF sample (n=8153). Training the same model on half of the Canadian general population sample (n=12518) and tested on the other half yielded higher accuracy (96.33%), and specificity (96.45%), but lower sensitivity (65.38%), specificity and AUC (0.9404). On the other hand, a conditional inference random forests model of suicide attempts, containing 4 variables, yielded an accuracy, sensitivity, specificity, and AUC of 91.78%, 66.96%, 91.89%, and respectively 0.8562, when trained on the entire CAF sample (n=8153), and tested on the entire CAF sample (n=25036). The same model, trained on half of the CAF sample (n=4077) and tested on the other half, misclassified all the suicide attempts cases, yielding an AUC of 0.5.

5.6.4 Overall results

Final conditional inference random forests models had between 2 and 9 predictors which were derived from antecedent best reduced models. Testing on unbalanced blind test sets yielded accuracies up to 99.71%, sensitivities up to 82.61%, and AUCs up to 0.9793, while specificity estimates mirrored the ones for predictive accuracies (Table 7.17).

Regardless of population setting and suicidal behaviour outcome, training the classifiers on larger training samples of different population domain than test samples yielded higher sensitivities on between-sample external validation. For example, training a model of suicidal ideation on the entire Canadian general population sample and testing on the entire CAF sample (Scenario 1) yielded a sensitivity of 79.54%, while training on the entire military sample and testing on the much larger civilian sample (Scenario 2) yielded a much lower sensitivity, that approached chance levels (51.45%). Similar difference in sensitivities were observed for suicide attempts (82.61% in Scenario 1 and 66.96% in Scenario 2), and suicide plans models, (58.27% in Scenario 1 and 58.24% in Scenario 2).

On the other hand, training the classifiers on same size samples from the same population domain yielded sensitivities of similar magnitude on within-sample external validation, regardless of population setting and suicidal behaviour outcome. The only exception was the model for suicide attempts in the CAF, in which all positive cases were misclassified, due to the very low prevalence of the outcome in this population. Indeed, models of suicidal ideation trained and tested on equally sized samples from the same original civilian and military samples yielded sensitivities of 64.10%, and 64.94% respectively. Likewise, models of suicide plans trained and tested on equally sized samples from the same original civilian and military samples yielded sensitivities of 68.97% and 67.16%, respectively.

5.7 Conditional prevalence of past 12-month suicidal behaviour outcomes in Canadian general population and Canadian Armed Forces

The conditional prevalence of past 12-month suicidal behaviour outcomes by population setting and type of construct within each risk algorithm is shown in Tables 7.23 and 7.24.

5.7.1 Suicidal ideation

The prevalence of past 12-month suicidal ideation in the Canadian general population ranged from 10.93%, in subjects unable to cut down alcohol use, to 76% in those experiencing all 9 components of the risk algorithm for suicidal ideation in the civilians. On the other hand, the prevalence of past 12-month suicidal ideation in the CAF ranged from 17.20%, in subjects who felt that needed help for mental health problem but did not receive it, to 36.36% in those experiencing all 5 components of the risk algorithm for suicidal ideation in the military.

5.7.2 Suicide plans

The prevalence of past 12-month suicide plans in Canadian general population ranges from 7.89%, in subjects who felt so sad that nothing could cheer up, to 53.37% in those experiencing all 6 components of the risk algorithm for suicide plans in the civilians. On the other hand, the prevalence of past 12-month suicide plans in the CAF ranged from 15.38%, in subjects with past-year major depressive episode significantly interfering with person's life, to 80% in those experiencing both components of the risk algorithm for suicide plans in the military.

5.7.3 Suicide attempts

The prevalence of past 12-month suicide attempts in Canadian general population ranges from 3.63%, in subjects receiving medication for their mental health problem in the past year, to 32.65% in those experiencing all 6 components of the risk algorithm for suicide attempts in the civilians. On the other hand, the prevalence of past 12-month suicide attempts in the CAF ranged from 1.96%, in subjects receiving treatment for own mental health problem in the past year, to 4.50% in those experiencing all 3 components of the risk algorithm for suicide attempts in the military.

5.7.4 Overall results

Experiencing all the components of a risk algorithm increased the prevalence of suicidal behaviour outcomes, regardless of population setting. Relative to the CAF, the conditional prevalence of suicidal ideation and suicide attempts was higher in the Canadian general population (76% vs. 36.36% and 32.65% vs. 4.50%, respectively).

6. Discussion

Suicidal thoughts, suicide plans, and attempts add to the public health burden of suicide, by their profound impact on family, friends, and communities. The worldwide recognition of suicide prevention as a public health priority has encouraged much research in this area during the past half century, with disappointing results: the mortality rate due to suicide and the rates of non-fatal suicide outcomes have not abated over the past decade, while predicting suicide behaviour seems not to be improving beyond chance levels [101,129]. Adopting new methodologies, such as machine learning algorithms, that allow for combining risk factors in ways that may increase the predictive accuracy of suicide behaviour is one promising avenue toward effective prevention and treatment [36,101,129,131].

This is the first study to use an ensemble recursive partitioning approach to identify risk algorithms of suicidal ideation, suicide plans, and attempts in the Canadian general population and Canadian Armed Forces.

As we have shown, many factors influence the predictive accuracy of a classifier, such as the availability of influential predictors for the problem at hand, the dimensionality of the data, the class imbalance, the number of cases available for training, the selection bias towards highly-correlated variables, and whether the algorithm has been optimally trained through a thorough selection of hyper-parameters values. We successfully optimized the algorithm at each step, and dealt with the dimensionality of the data during the model selection stage, by reducing the size of both the feature set, and the learning set through down-sampling, while simultaneously balancing the outcome classes. To mitigate the selection bias towards highly-correlated variables, we employed a *conditional permutation accuracy* scheme that allowed for an appropriate final model selection [185]. Finally, we used unbalanced validation and blind testing sets to ensure that the predictions reflect the probability of suicidal behaviour outcomes in the real world [153] (Figure 4.1). Thus, the proper use of conditional inference random forests methodology enabled us to identify concise models that conceptualize suicide behaviour as resulting from cross-sectional risk factors that interact synergistically, with satisfactory predictive performance.

6.1 Summary of results

In brief, we showed that, regardless of population setting:

1. partial hyper-parameter tuning of the conditional inference random forests algorithm improved the model's performance;
2. model reduction through sequential variable selection resulted in an overall increase in predictive accuracy;
3. only a handful of variables were important in predicting suicidal behaviour; among them, (i) past 12-month suicidal ideation, (ii) past 12-month mental health treatment, (iii) past 12-month MDE interfering with the person's life, (iv) past 12-month unmet need for help with mental health problems and lifetime indicators of (v) death ideation, (vi) worry, and (vii) problem drinking ended up in more than one final risk algorithm;
4. each model's performance on within-sample external validation was satisfactory, with moderate to high sensitivity and AUC;
5. each model's performance on between-sample external validation was conditional on the size and heterogeneity of the training sample, larger, more diverse samples yielding better sensitivities than lower, more homogeneous samples; if trained on smaller, more

homogenous samples, performance of a model is likely to worsen on between-sample external validation than on within-sample external validation;

6. all models of suicide attempts performed better during external validation than those of suicidal ideation;
7. relative to individuals experiencing only one influential risk factor, the conditional prevalence of suicidal behaviour outcomes in those fulfilling all components of any risk algorithm is several folds higher.

6.2 Partial hyper-parameter tuning

Regardless of outcome and population setting, increasing the default value of the number of variables to be randomly chosen at each split (*mtry*) resulted in an absolute increase of the predictive accuracy of the full models. On the other hand, conditional inference random forests were relatively robust to the settings of the *ntree* tuning parameter, confirming that training on larger forests does not always yield better results [154]. In accord with previous research [204], we showed that an increase of the default number of trees to be grown in the forest did not translate into an improvement in predictive accuracy. Our results were also in line with previous research showing that *mtry* is the primary tuning parameter in random forests, having the greatest impact on the complexity of the model and its predictive accuracy [164,205-207]. We agree with Goldstein et al. [207] that even coarse tuning warrants an optimal choice of *mtry* as relying on its default value does not guarantee the best results.

6.3 Predictive accuracy during model selection

Regardless of outcome and population setting, model reduction through recursive elimination of non- and least-informative predictors yielded models with improved predictive performance over previous larger models. All final models, which were based on subsets of predictors representing much smaller feature spaces than the original ones, performed better than their corresponding best full models. Our results are consistent with previous research showing the efficiency of recursive feature elimination in selecting small subsets of influential variables with good predictive accuracy [208]. We do not concur with previously reported opinions that accurate prediction of suicide behaviour may require combination of hundreds of risk factors [137].

6.4 Feature space reduction through variable selection

Regardless of outcome and population setting, the proportion of non-important predictors was several folds larger relative to the important ones. Indeed, the final subsets were minute fractions of the initial original sets of predictors, supporting our Hypothesis 2a that (1) the majority of predictors were not relevant for the suicidal behaviour outcomes, and (2) only a small number of influential predictors will be selected in the final risk algorithms.

6.5 External validation

With one exception, we obtained high overall accuracies, specificities and AUCs for all models, regardless of population setting. However, these metrics may be misleading in situations in which the outcome is rare: an algorithm may perform well in terms of overall accuracy, specificity, or AUC, because it can successfully identify controls, but perform poorly in terms of identifying cases [154,156,158]. When the outcome is as rare and potentially life threatening as suicide attempts, the focus should be on sensitivity over specificity, given the dire consequences for the occurrence of false negative cases [137]. With one exception, all models had fair to moderately high sensitivities despite being carefully calibrated to reflect the probability of the outcome in the real world [153]. This suggests that missing important risk factors rather than methodological deficiencies are responsible for the lower performance in some of the models. For example, models of suicidal ideation and suicide attempts that have been trained on the larger, more diverse civilian samples, are better than others in capturing suicidal behaviour cases. On the contrary, true positives are less likely to be captured by the current models of suicide plans, regardless of population setting, or by any all models trained on smaller, more homogenous military samples. The differences in between-sample and within-sample predictive performance are addressed below.

6.5.1 Within-sample external validation

Regardless of outcome and population setting, within-sample external validation (i.e. training and testing the algorithm on equal-sized samples of the same population domain) yielded sensitivities within a tight range (64.10% to 68.97%), suggesting that the final algorithms miss some important suicidal behaviour risk factors which were assessed in neither of the current surveys. The present risk algorithms were built on constructs from only a handful of domains related to mental health treatment and experiences, depression, suicide, GAD, and AUD. We have no doubt that additional constructs, related to psychopathology, negative life events, social connectedness or social support, may have further improved the performance of the current models, if available for assessment.

6.5.2 Between-sample external validation

Regardless of outcome, algorithms trained on more predictors and the larger, more diverse civilian samples yielded better sensitivities during between-sample external validation (i.e. training and testing the algorithm on full size samples of different population domains) than those trained on less predictors and smaller, more homogenous military samples: 79.54% vs.

51.45% for suicidal ideation, 58.27% vs. 58.24% for suicide plans, and 82.61% vs. 66.96% for suicide attempts, respectively.

6.5.3 Comparison between within-sample and between-sample external validation

We obtain lower sensitivities on between-sample validation than within-sample validation for models of suicidal ideation and suicide attempts trained on military samples, as well as models of suicide plans, regardless of training sample, partially supporting Hypothesis 3b (i.e. an algorithm applied on a civilian sample will perform less accurate in predicting any of the three outcomes in military personnel). Our results are consistent with a recent study evaluating 127 prediction models that found that the predictive performance may worsen substantially on external validation [144]. We hypothesized that a model's predictive performance will degrade when the training and testing are performed on different samples than on samples from the same population setting. Predictive performance is intimately related to the degree of similarities between the training and testing sets. Usually, the performance of an algorithm degrades when applied to samples that differ significantly from the ones on which it was developed. This is due mainly to underfitting, which occurs when the training sample lacks important predictors of the outcome [209,210]. Thus, we expected that any difference in performance between the civilian and military testing samples was accounted for not only by the size and the heterogeneity of the training samples, but by whether the risk algorithm on which the model was trained missed important predictors of the outcome that were more specific to the testing sample. Applying models trained on the smaller, more homogenous military sample to the larger, more heterogeneous civilian sample yielded lower sensitivities than when the training and testing were done on military samples half the size of the original CFMHS sample. This was less likely due to the difference in the size of the training samples and the number of available positive cases, but more to the fact that the risk algorithms for suicidal ideation and attempts in military missed important risk factors that played a lesser role in Canadian Armed Forces than in Canadian civilians, such as unemployment, lack of social integration and support, or poorer health outcomes. Indeed, active military personnel not only have a distinct socio-demographic composition relative to their civilian counterparts, but they are a distinct component of the labour force, that is predominantly male, full-time employed and fairly well remunerated (Table 7.22)[8]. Moreover, they are generally healthier relative to civilians in the same age group range, due to the rigorous health and fitness standards required at recruitment and ongoing during the military service [143], and benefit from an increased sense of social connectedness due to a strong sense of community and camaraderie among peers who share the same professional ethic, ethos, and value system [20,38]. On the other hand, military personnel are exposed to unique stressors, such as deployment-related occupational trauma that occurs in addition to other sources of lifetime trauma to which any of their civilian counterparts is exposed [8].

As for the models of suicidal ideation and suicide attempts in civilians, we obtained higher sensitivities on between-sample validation (79.54% and 82.61%, respectively) than within-sample validation (64.10%, and 65.38%, respectively), apparently contradicting Hypothesis 3b. This finding is likely explained by the stark differences in size of the civilian samples available for training. A model performs poorer during validation if (1) a sizeable part of the original sample size is not offered for training, and (2) the risk algorithm used for training misses important predictors of the outcome. We did not observe a degradation of performance on between-sample validation. This suggests that the predictors for suicide ideation and attempts are more likely to be shared between Canadian civilians and military personnel, than to be specific to either population setting.

Sensitivities on between-sample external validation of models of suicide ideation and attempts trained on civilian samples were the highest among all models, suggesting that the true positives are likely to be captured by this method.

6.5.4 Comparison between external validation of models of suicidal ideation and suicide attempts

Regardless of the population setting, all models of suicide attempts performed better during external validation than those on suicidal ideation, fully supporting Hypothesis 2b, which hypothesized that predictive accuracy vary by outcome type: models of suicide attempts would have higher accuracy than those of suicidal ideation, because more severe outcomes are associated with more discriminative predictors and the generalizability of a model is contingent on how discriminative its features are with regards with the outcome. Our results showed that virtually all top predictors of suicide attempts had higher mean decrease in predictive accuracy than the corresponding predictors of suicidal ideation (Tables 7.5 to 7.10).

6.6 Final risk algorithms and suicidal behaviour outcomes prevalence

We used conditional inference random forests with a conditional permutation accuracy scheme to explore predictive importance of potential risk factors of suicidal behaviour in Canadian general population and the Canadian Armed Forces. Relative to people experiencing only one of the influential risk factors, conditional prevalence of each suicidal behaviour outcome increased several folds in those experiencing all components of the risk algorithm (Tables 7.23 and 7.24).

Some predictors, which were important in one population sample played a lesser role in the other and did not end in the final risk algorithms. This supports our hypothesis that models will vary due to the different structure of the two populations (Hypothesis 3a).

6.7 Common predictors across risk algorithms

The predictors identified in the present study worked synergistically to maximize model's performance. Therefore, we caution against any attempt to interpret influential predictors in isolation, because their importance was determined by considering the risk algorithm as a whole, with all its components interacting simultaneously [136,137]. We agree with Walsh et al. that any interpretation of the results should focus rather on the risk algorithm as a whole and broad patterns of importance than on isolated components and narrow, very specific conclusions [137]. With this cautionary remark in mind, we found that some predictors consistently ranked higher during model selection and ended up in more than one final risk algorithm, regardless of the population setting. This suggests that these predictors, when jointly interacting with other influential risk factors, are influential on predicting suicidal behaviour.

6.7.1 Passive suicide ideation

Passive suicidal ideation (*I am better off dead*), in the context of interacting with other influential risk factors (Tables 7.18 to 7.21), was one of the strongest predictors of suicidal ideation, suicide plans and attempts in civilians. This is in line with previous research showing that passive suicide ideation can be a marker for suicidal behaviour [211]. However, passive suicidal ideation was absent from any model trained on military samples, suggesting that the processes underlying active suicidal ideation are different for civilian and military personnel. Interestingly, passive suicide ideation was associated with active suicidal ideation (*I want to kill myself*) in the absence of hopelessness (*This will never change*). Hopelessness did not rank high enough among the most important predictors for suicide ideation to be chosen in the final risk algorithms. This finding contradicts one of the main assumptions of Joiner's Interpersonal Theory of Suicide, which states that hopelessness is the main driver of the transition from passive to active suicide ideation [21]. One possible explanation is that the recall period for being hopeless about the future was lifetime instead of a shorter period of time prior to the interview. As such, hopelessness was a poor predictive factor, even in combination, because the recall period was long. Few would expect that being hopeless several years prior to assessment would predict past-year suicide ideation. On the contrary, hopelessness would have been a strong predictor if asked over a much shorter period of time, such as in an interviewed bereaving individual with low social support [101].

6.7.2 Past 12-month active suicidal ideation

Past 12-month active suicidal ideation (*I want to kill myself*), in the context of interacting with other influential risk factors (Tables 7.19 to 7.21), was one of the strongest predictors of past 12-month suicide plans and attempts in civilians and suicide plans in military. Our findings are consistent with previous research showing that suicidal ideation predicts suicide attempts and suicides in both civilian [211,212] and military samples [15,50,135,136]. We notice the presence of past 12-month suicide ideation among the most influential risk factors of past 12-month suicide attempts in all but the last reduced model of suicide attempts in the military. In our opinion, this is less likely an indication that processes underlying suicide attempts may be different for civilian and the military, but a consequence of the data-driven feature reduction process, in which a variable is kept in the subsequent reduced models only if its standard deviation of importance is greater than a threshold value that is conditioned on the variability of least-important predictors in the model being assessed [165,202].

6.7.3 Past 12-month treatment for mental health problems

Indicators of past 12-month mental health treatment, including medication, in the context of interacting with other influential risk factors, were among of the strongest predictors of past 12-month suicidal behaviour, regardless of outcome and population setting (Tables 7.18 to 7.21). This is consistent with recent research using machine learning algorithms showing that medication and health care utilization is predictive of future suicides [135,136] and suicide attempts [137]. However, our findings should be interpreted with caution, as mental health service use is associated with severity of illness and comorbidity [213-215], which are themselves powerful determinants of suicidality [25,26,29,31,32,37,42,216]. Therefore, it is plausible that indicators of treatment identified in the present research, including medication, may act in fact as proxies for latent variables of care or missing psychiatric diagnoses [137]. We concur with the later possibility as health care utilization and perceived need for care were extensively assessed in the present mental health surveys, even though very few of the many variables derived from these modules ended up as influential predictors of suicidal behaviour.

6.7.4 Past 12-month unmet need for help with own mental health problems

Past 12-month unmet need for help with own mental health problems, in the context of interacting with other influential risk factors, was one of the strongest predictors of past 12-month suicidal behaviour in civilians (Tables 7.18 to 7.20). However, it was not included in the final algorithms of more severe outcomes in the military (Table 7.21), suggesting that the lack of unmet need for help in the Canadian Armed Forces may be accounted for by the differences in access to mental health care across the two populations. Military personnel access a separate mental health system than civilians, that offers a variety of clinical services, including mandatory periodic health evaluations, mandatory pre- and post-deployment mental health screening, as well as clinical mental health care through several clinical programs [217], with presumably less barriers to care than the Canadian civilian mental health system. This view is consistent with recent research showing that perceived sufficiency of mental health care in Canadian Armed Forces is higher than in full time employed civilian counterparts with similar socio-demographic characteristics [218].

6.7.5 Past 12-month MDE with interference on person's life

One finding that stands aside from previous research on risk factors for suicide behaviour is the fact that the severity of major depressive episode, in combination with other cross-sectional risk factors, predicts suicide plans and attempts in ideators. Previous research on ideators found that a diagnosis of major depression is a poor predictor of the transition from ideation to plans and attempts [9,36,37]. However, none of those studies consider the severity of depression, in terms of its impact on person's daily activities and responsibilities. Impairment in one or more areas of functioning belongs to the concept of clinical significance of symptoms and syndromes [188]. This concept is part of the diagnostic criteria of many mental disorders, including MDE, and it is related to further concepts of symptom prominence and severity [188,219]. We found that past-year MDE, in combination with other risk factors, predicts past-year suicide plans in both civilian and military past-year ideators as well as past-year suicide attempts in past-year civilian ideators, only when it significantly interferes with normal routine, occupational functioning, social activities or relationships. The same risk factor, while absent from the final risk algorithm for suicide attempts in the military, was present among the top 10 most influential risk factors up to the second to last model.

6.7.6 Worry

Interestingly, a worry item (being worried about own mental health) was part of the final risk algorithms of suicidal ideation in both civilian and military samples as well as of the final model of suicide attempts in the Canadian Armed Forces. Moreover, worry items (being worried about own mental health, about relationships with family, or about love life) were present in all of the antecedent reduced models of suicidal behavior, regardless of population setting. Excessive and unrealistic worrying is one of the hallmark features of GAD [188]. However, a GAD diagnosis did not qualify as a component of any civilian and military final risk algorithms. Moreover, a recent meta-analysis found that anxiety-specific constructs were weak predictors of suicidal behavior when analyzed in isolation [139]. Furthermore, the exclusive association between worry and GAD is currently debated. Worry has been found to be common in other psychological disorders [220], some of which, such PTSD, being stronger predictors of suicidal behaviour than GAD [36,37]. Unfortunately, PTSD was only assessed in the CFMHS survey and its constructs were not included in the present analyses. It is therefore plausible that the worry items we found in the present analyses were simply proxies of PTSD or another mental health disorder in which worry is present to a meaningful extent, and not exclusive GAD-related constructs. However, our results are indicative of the predictive potential of worry items when analyzed in conjunction with other, potentially stronger, suicide-related risk factors, such as mental health care utilization and psychopathology, suggesting that comorbidity of symptoms may be an important construct in predicting suicidal behaviour.

6.7.7 Problem drinking

Two items related to alcohol dependence in civilians (unable to cut down alcohol use despite making several attempts to quit) and alcohol abuse in military (drinking alcohol frequently interfered with work or responsibilities), in the context of interacting with other influential risk factors (Tables 7.20 and 7.21), were among the strongest predictors of suicidal ideation. These findings are consistent with previous studies showing that heavy or binge drinking and alcohol-related problems are independent predictors of suicide in civilians [19] and military personnel [221,222]. Alcohol has been coined the lubricant to suicidality [223] as it may increase the suicide risk via complex effects and actions, such as greater impulsiveness, impaired problem solving, and depressant effects through its impact on brain neurophysiology [222,223].

6.7.8 Excluded risk factors and domains

All influential predictors identified in the present risk algorithms were derived from only half of the available modules. No socio-demographic characteristic or item pertaining to work stress, socio-provision scale, or childhood abuse emerged as influential predictor in our civilian and military models. We hypothesized that proxy constructs of thwarted belongingness (living alone, low sense of belonging to local community, low score on Social Provision Scale) and perceived burdensomeness (lost self-confidence, worthlessness, guilt) will be consistently informative across model selection process and rank high among all the candidate predictors for suicidal ideation (Hypothesis 1a). Moreover, we hypothesized that child abuse items will emerge as strong predictors in all models (Hypothesis 1b). We found instead that all these constructs were weak predictors of suicidal behaviour outcomes across model development. Our results are in line with recent research using machine learning algorithms on military samples finding that models of suicide after outpatient mental health visits were lacking socio-demographic and Army career characteristics indicating disadvantaged social or career status [136]. On the other hand, socio-demographic characteristics, such as age, sex, and race were found to be consistently informative in terms of predicting suicide risk in models using machine learning methodology in civilian samples [137]. One explanation for the missing hypothesized predictors from the final risk algorithms is that stronger predictors may have mediated the effects of weaker predictors [136]. Take for example childhood trauma. There is strong evidence from population-based studies that child abuse is an important determinant of both adult suicidal behaviour [68-70,72,74-76,224,225] and mental health problems, including depression [74,75,226]. Thus, it is plausible that part of the association between child abuse and suicidality may be mediated by psychopathology with onset closer to the suicidality outcomes being measured [224,225].

6.8 Comparison with other findings

Comparing our results to previous research using machine learning for predicting suicidal behaviour outcomes is not straightforward, due to the differences in methods employed, type and size of population domains, and feature spaces used for training the algorithms. However, in general, we obtained similar predictive performance or better than most of the previous studies. For instance, studies using longitudinal electronic health records to predict suicidal behaviour had AUC ranging from 0.76 to 0.84 [137,227], while studies focused on predicting attempts on more restricted population domains, such as patients with psychiatric diagnosis, had AUC estimates between 0.65 and 0.87 [228,229]. Furthermore, two studies focused on predicting suicide in military samples found AUCs ranging from 0.67 to 0.84 [135,136]. Our results are also in line with previous research focused on identifying the most important predictors for suicidal behaviour: indicators of suicidality [132,135,136], psychiatric diagnosis

[132,136,137,227], and psychiatric treatment [135-137] were among the most influential predictors of fatal and non-fatal suicide attempts in both civilian and military samples.

6.9 Limitations

This research has several important limitations related to the nature of the data available for analysis and the particularities of the methodology used in the present research.

6.9.1 Data-driven limitations

First, due to the cross-sectional nature of the data, we could not establish any temporal relationship between the risk algorithm identified in the present research and the suicidal behaviour outcomes. Same data-driven limitation precluded us from evaluating the temporal validation of the models within the same population setting.

Second, the cross-sectional design of the mental health surveys did not approximate the inner processes of Interpersonal Theory of Suicide, in which distal risk factors, combining with proximal ones over relatively short periods of time, contribute to the advent of a suicide outcome and to the transition to a more serious one [21,100]. Moreover, no candidate predictor was assessed for the purpose of informing any of theory's major constructs: thwarted belongingness, perceived burdensomeness, and acquired capability for suicide. This may explain why many of the risk factors stipulated by the theory did not rank high in our models. In addition, one recent systematic review aimed at identifying support for theory's main predictions found that the three-way interaction between these constructs was less straightforward than originally stated in predicting suicide attempts [230]. The authors noticed also that many of the studies used to assess Interpersonal Theory of Suicide relied on cross-sectional designs, which may explain the mixed evidence found in their review.

Third, the military sample was not large enough to allow for a reliable prediction of suicide attempts in the CAF. For most outcomes, we were able to split each dataset into training and testing sets that permitted us to perform proper model validation. However, the number of military suicide attempts was so low that we had to use the entire CAF sample for training and testing, being in the impossibility of splitting the data using the hold-out approach. We had to rely on a different resampling scheme (i.e. *k*-fold cross validation) to properly train the algorithm in order to obtain satisfactory performance metrics on between-sample external validation. However, this approach did not help with within-sample external validation, because the numbers of cases available for training was even lower, about half of the initial sample. The poor performance we observed in this instance is most likely a consequence of the way most machine learning algorithms, including the one we used in the present research, are built: to minimize overall misclassification error [151,152]. The model preferred to err on the side of caution and classify every case in the testing set as a non-attempter (Table 7.17a and 7.17b), indicating the limits of the method when the number of available cases is too low (i.e. less than 20).

Fourth, we were missing other influential variables that were not available for both datasets. For example, addition of traumatic experiences, precipitating stressful life events, chronic pain, and co-morbid mental disorders, especially those known to be associated with a high risk for suicide [53-59,135,136] may have further improved the predictive accuracy of suicide attempts with the current models. Unfortunately, only those constructs that were assessed in both surveys were included in the present analysis. Traumatic experiences, PTSD and panic disorder have been assessed only in the CAF sample, precluding us from including them as candidate predictors.

These predictors are likely to be associated with suicidal behaviour outcomes [36,37,52-61,69-76,137] and have differential distributions across samples. We also missed the opportunity to explore any potentially protective factors, such as optimism, which has been shown to be associated with less severe suicide ideation, by buffering the effects of hopelessness in military personnel [231]. While both surveys assessed emotional well-being and positive functioning as part of the Positive Mental Health module, the recall period was shorter than the one for the outcome, therefore constructs derived from these modules were not included in the present analysis.

6.9.2 Algorithm-related limitations

Conditional inference random forests algorithm is a computationally intensive method. With limited computer resources available, we were forced to adopt several trade-off strategies to reduce computational processing time, while preventing the algorithm from exhausting the available physical memory. How this approach affected the selection and the performance of the final models is unknown. It is entirely possible that employing the same methodology, in the absence of the any restrictions on computational resources may have produced more reliable models of suicidal behaviour, with better predictive accuracy on external validation.

First, to reduce computation time during model selection, we balanced the training set by under-sampling the majority class so that its frequency equals that of the minority class. However, this approach has a major drawback, when dealing with extremely imbalanced datasets, such as ours: it drastically reduces the number of training instances, due to the scarcity of the suicide-related outcomes. Alternative strategies, relying on combining different resampling approaches might have selected different sets of optimal predictors, with better predictive performance. For example, strategies combining under-sampling and over-sampling have been shown to be effective in solving classification problems from imbalanced data [159].

Second, to obtain reliable estimates while preserving computation time, our strategy was to perform feature selection sequentially, by using the model with lowest hyper-parameter values, for which the predictive accuracy was similar with that of the best performing model. We consequently relied on less dense random forests, of 500 trees each, to generate permutation VIMs. This approach may have yielded less reliable importance scores. Moreover, feature selection optimization produced optimal reduced models with a predictive accuracy within 1 standard error or higher than either the full model or the first reduced model, without non-informative predictors. However, this does not guarantee that there are no other optimal models among all the possible candidate classifiers.

Third, we did not examine how variable rankings were affected by the choice of the hyper-parameters values. It is well known that not only the predictive accuracy is affected by this choice, but also the VIMs [207]. Variable selection through feature space reduction benefited from a partial tuning grid-based examination of *mtry* values during each step of the model selection process. However, to preserve the computational efficiency during the variable selection step, we fixed the value of *ntree* to its default after hyper-parameter optimization of the full models. How different values of *ntree* could have affected the variable rankings in subsequent reduced models is unknown. Studies have shown that larger forests lead to more refined and stable variable importance scores, which eventually translates into more reliable feature space reduction and consequently better choice of the final risk algorithms [205,207]. Optimizing the algorithm at each step of the algorithm by employing a tuning grid with more granularity in term of *mtry* and *ntree* hyper-parameters might have yielded other optimal feature subsets and subsequent models with better predictive performance.

Fourth, the algorithm used in the present research was not designed to allow for the possibility of dealing with complex survey data. Therefore, our results are not representative for the populations from which these samples were drawn.

Lastly, we did not include an alternative conventional method for comparison because using parametric methods is unfeasible when dealing with large feature spaces, as the modelling of a staggering number of high-order interactions is not possible. Additionally, we would have run into the large *p*, small *n* problem (that is, more variables than available observations), due to the use of subsampling of the majority class during resampling. Moreover, we have been persuaded by the recent research on the matter that such addition would add little to the current body of knowledge regarding the limitations of the traditional approaches in predicting suicidal behaviour [101,129,137].

6.10 Strengths

Our study has several notable strengths that are worth mentioning. We identified subsets of influential predictors of suicidal behaviour by (1) performing complex variable selection through the incorporation of high order interaction effects and nonlinearities impossible to account for with traditional methodologies, such as logistic regression, while concomitantly (2) attenuating the correlation between variables, (3) controlling for overfitting, and (4) mitigating the severe imbalance of the data through undersampling during resampling, thus ensuring that the variation generated by undersampling is properly captured [153]. Moreover, the models were calibrated to reflect the prevalence of the outcomes in the real world: while training sets were balanced, those used for internal and external validation were kept unbalanced. Furthermore, the feature selection process was entirely data-driven [165] and not subject to arbitrary strategies [164] for the estimation of variable importance thresholds. In addition, all the results of the present research included an understanding of the effects of the method's adjustable hyper-parameters. Indeed, variable selection through feature space reduction benefited from a tuning grid-based examination of *mtry* values during each step of the model selection process. Previous studies have shown that an optimal choice of *mtry* is at least as important for model performance as is for variable selection [205]. With one notable exception, the original sample sizes were sufficiently large to put aside test datasets for external validation. It has been suggested that internal validation on subjects within the same population domain as the one for which the algorithm was trained, even in subjects recruited at a later point in time, may not suffice before a prognostic model can be applied in practice [232]. Regardless of population setting and outcome, we were able to test the final models of suicide behaviour on subjects from different population settings.

6.11 Future areas of research

Our research is an important step forward in terms of developing suicide risk algorithms based on large national-representative samples. However, several aspects of our approach could be improved in future research. First, comparing the conditional permutation accuracy scheme with classical approaches for variable importance measures may shed light into how highly correlated predictors influence the construction of suicide risk algorithms and the performance of predictive models [185]. Second, maximizing available training instances in extremely imbalanced datasets such as those involving suicidal behaviour is critical for selecting risk algorithms with optimal predictive performance. Future studies should rely on calibration techniques capable of blending different resampling approaches, such as synthetic minority over-sampling, which uses a combination of under-sampling the majority class and over-sampling the minority class [159]. Third, future research using machine learning algorithms should compensate for complex survey designs to ensure that the model built from survey data is representative of the population. Machine learning studies incorporating survey weights are few [233] and most of the algorithms, including the one used in the present research, were not designed to allow for the possibility of dealing with complex survey data. Fourth, future work should address the potential differences in the processes underlying suicidal behaviour in civilian and military personnel, by focusing on more calibrated civilian and military samples. Studies focusing on understanding the potential effects of military service on suicidal behaviour outcomes require truly comparable non-military samples [8]. Finally, future research may benefit from investigating whether other machine learning algorithms perform better than the methodology we used, while fully accounting for the critical aspects currently identified, such as the complex survey design, the severe imbalance of the data, and the high correlation between predictors.

6.12 Conclusion

Our research has shown that employing conditional inference random forest methodology on data derived from large nationally representative mental health surveys has the potential of generating models of suicidal ideation, suicide plans and attempts that not only reflect the complex nature of suicidal behaviour in both civilian and military population settings, but indicate that the true positive cases are likely to be captured by this approach. Moreover, the high concentration of suicidal behaviour outcomes in individuals experiencing all items within identified risk algorithms justifies prospective validation studies in at-risk populations that may inform suicide prevention strategies in both civilian and military populations.

7. Tables and graphs

Table 7.1. Hyper-parameter tuned predictive performance of full[1] conditional random forest models by population setting and suicidal behaviour outcome

ntree	Hyper-parameters: *Mtry*									
	√p	2√p	4√p	8√p	16√p	√p	2√p	4√p	8√p	16√p
	Predictive accuracy[2] %(SE)									
	Suicidal Ideation									
	Canadian General Population (n_{train}=12519, p=290)					**Canadian Armed Forces** (n_{train}=4077, p=289)				
500	74.75(0.10)	76.50(0.12)	77.50(0.12)‡	77.82(0.14)†	77.75(0.14)	73.15(0.21)	75.53(0.24)	76.86(0.21)	77.61(0.27)	77.25(0.27)
1000	74.77(0.10)	76.52(0.10)	77.55(0.11)	77.80(0.14)	77.65(0.15)	73.12(0.19)	75.54(0.24)	76.97(0.24)	77.66(0.27)	77.24(0.25)
2000	74.82(0.09)	76.46(0.10)	77.57(0.10)	77.77(0.13)	77.70(0.14)	73.20(0.17)	75.51(0.24)	76.97(0.23)	77.74(0.27)†	77.16(0.27)
	Suicide Plans									
	Canadian General Population (n_{train}=12518, p=294)					**Canadian Armed Forces** (n_{train}=4076,p=293)				
500	80.73(0.16)	82.10(0.16)	84.69(0.14)	93.29(0.21)‡	94.25(0.14)†	79.15(0.22)	79.63(0.25)	80.56(0.27)	83.57(0.35)	86.30(0.45)
1000	80.88(0.16)	82.11(0.15)	84.79(0.15)	93.36(0.22)	94.25(0.14)†	79.07(0.23)	79.59(0.28)	80.65(0.29)	83.52(0.32)	86.35(0.44)
2000	80.83(0.16)	82.15(0.14)	84.86(0.16)	93.33(0.21)	94.23(0.14)	79.04(0.19)	79.46(0.21)	80.63(0.26)	83.59(0.32)	86.46(0.43)†
	Suicide Attempts									
	Canadian General Population (n_{train}=12518, p=298)					**Canadian Armed Forces** (n_{train}=8153, p=297)				
500	80.39(0.20)	80.89(0.17)	84.56(0.35)‡	86.40(0.26)	84.96(0.37)	76.57(0.43)	77.96(0.52)	78.78(0.68)	80.48(0.69)	80.65(0.90)
1000	80.40(0.22)	80.84(0.18)	84.88(0.33)	86.47(0.25)	85.04(0.37)	76.26(0.44)	77.61(0.52)	78.83(0.62)	80.57(0.69)	80.74(0.91)
2000	80.56(0.19)	80.95(0.15)	84.84(0.34)	86.61(0.27)†	85.02(0.37)	76.35(0.45)	77.52(0.53)	78.78(0.68)	80.26(0.73)	80.76(0.89)†

[1]Including all p predictors; [2]Averaged over 50 iterations; †Best full model is the model with the highest predictive accuracy; ‡Optimal full model is the least performant full model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best full model cannot be rejected; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown in a forest; p= number of variables in the model; n_{train}=number of observations used for training the classifier.

Table 7.2. Corrected resampled t-tests[1] comparing hyper-parameter tuned predictive performance of best[†] full[2] conditional random forest model and all antecedent[3] full[2] models, by population setting and suicidal behaviour outcome.

	Hyper-parameters									
ntree	*mtry*									
	√p	2√p	4√p	8√p	16√p	√p	2√p	4√p	8√p	16√p
	t-score									
	Suicidal Ideation									
	Canadian General Population (n_{train}=12519, p=290)					**Canadian Armed Forces** (n_{train}=4077, p=289)				
500	4.42***	2.11*	0.61‡	Reference†	-	4.48***	2.43*	1.31‡	0.33	0.59
1000	-	-	-	-	-	4.47***	2.41*	1.19	0.28	0.60
2000	-	-	-	-	-	4.35***	2.58*	1.18	Reference†	-
	Suicide Plans									
	Canadian General Population (n_{train}=12518, p=294)					**Canadian Armed Forces** (n_{train}=4076, p=293)				
500	14.11***	13.27***	10.20***	0.92‡	Reference†	2.88**	2.53*	2.09	0.89‡	0.24
1000	-	-	-	-	-	2.96**	2.80**	2.00	0.96	0.23
2000	-	-	-	-	-	3.14**	2.79**	2.02*	0.92	Reference†
	Suicide Attempts									
	Canadian General Population (n_{train}=12518, p=298)					**Canadian Armed Forces** (n_{train}=8153, p=297)				
500	4.57***	4.29***	1.21‡	0.37	1.05	1.07‡	0.75	0.55	0.12	0.07
1000	4.71***	4.33***	1.12	0.28	0.99	1.11	0.86	0.56	0.08	0.01
2000	4.64***	4.29***	1.04	Reference†	-	1.09	0.91	0.57	0.17	Reference†

[1]Two-tail, df=49, α=0.05; [†]Model with the highest predictive accuracy; [2]Including all p predictors; [3]Model for which the values of *mtry* and/or *ntree* are lower than the corresponding values of the best full model; [‡]Optimal full model is the least performant full model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best full model cannot be rejected; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown in a forest; p= number of variables in the model; n_{train}=number of observations used for training the classifier; *P< 0.05; **P< 0.01; ***P< 0.001.

Table 7.3. Predictive performance of reduced[1] conditional random forests models of suicidal behaviour in Canadian general population.

p	*mtry* √p	2√p	4√p	8√p	16√p	√p	2√p	4√p	8√p	16√p
	Predictive Accuracy[2] **%(SE)**					**t-score**[3]				
Suicidal Ideation (n_{train}=6260, *ntree*=500)										
290	74.56(0.16)	76.68(0.15)	78.06(0.20)‡	78.79(0.21)	79.00(0.20)†	4.45***	3.04**	1.2‡	0.32	Reference†
157	75.90(0.15)	77.22(0.18)‡	78.10(0.19)	78.51(0.22)†	-	3.06**	1.65‡	0.61	Reference†	-
115	76.73(0.15)‡	77.52(0.17)	78.21(0.16)	78.44(0.19)†	-	1.92‡	1.21	0.44	Reference†	-
81	77.60(0.16)‡	78.36(0.18)	78.50(0.19)	78.60(0.21)†	-	1.06‡	0.32	0.17	Reference†	-
60	78.26(0.14)‡	78.59(0.15)	78.73(0.17)	78.94(0.21)†	-	0.76‡	0.47	0.34	Reference†	-
50	78.64(0.17)‡	79.06(0.17)	79.13(0.18)†	-	-	0.60‡	0.12	Reference†	-	-
46	78.52(0.16)‡	78.90(0.17)	79.05(0.19)†	-	-	0.44‡	0.27	Reference†	-	-
36	78.96(0.17)‡	79.18(0.18)	79.40(0.19)†	-	-	0.66‡	0.42	Reference†	-	-
27	79.04(0.15)‡	79.46(0.18)	79.59(0.21)†	-	-	0.72‡	0.24	Reference†	-	-
22	78.83(0.15)‡	79.60(0.16)	79.83(0.19)†	-	-	1.52‡	0.46	Reference†	-	-
14	79.23(0.18)‡	80.26(0.21)	80.32(0.20)†	-	-	1.65‡	0.09	Reference†	-	-
4	78.00(0.24)	-	-	-	-	-	-	-	-	-
Suicide plans (n_{train}=12518, *ntree*=500)										
125	81.87(0.18)	83.99(0.17)	87.98(0.26)	94.31(0.12) †‡	-	14.00***	12.25***	5.99***	Reference†‡	-
53	84.02(0.15)	86.95(0.20)	94.28(0.27) †‡	-	-	8.05***	5.58***	Reference†‡	-	-
34	84.58(0.16)	88.55(0.24)	94.64(0.14) †‡	-	-	14.42***	7.26***	Reference†‡	-	-
24	86.02(0.16)	92.62(0.22)	94.39(0.12)	-	-	-	-	-	-	-
Suicide Attempts (n_{train}=12518, *ntree*=500)										
99	80.75(0.16)	82.49(0.28)	86.14(0.28) †‡	84.61(0.39)	-	4.17***	2.59*	Reference†‡	-	-
57	81.18(0.19)	84.60(0.31)	85.23(0.28)	84.54(0.38)	-	-	-	-	-	-

[1]Including a subset of cross-sectional predictors; [2]Averaged over 50 iterations; [3]Corrected resampled t-tests (two-tail, df=49, α=0.05) comparing hyper-parameter tuned predictive performance of best† conditional random forest reduced model and all antecedent[4] reduced models; †Model with the highest predictive accuracy; [4]Model for which the values of *mtry* and/or *ntree* are lower than the corresponding values of the best reduced model; ; ‡Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown in a forest; p= number of variables in the model; n_{train}=number of observations used for training the classifier;
*P< 0.05;**P< 0.01;***P< 0.001.

Table 7.4. Predictive performance of reduced[1] conditional random forests models of suicidal behaviour in Canadian Armed Forces.

	mtry							
p	√p	2√p	4√p	8√p	√p	2√p	4√p	8√p
	Predictive Accuracy[2] %(SE)				t-score[3]			
Suicidal Ideation (n_{train}=4077, *ntree*=500)								
141	74.75(0.20)	76.17(0.22)ǂ	77.24(0.25)	77.40(0.25)	3.04**	1.55ǂ	0.21	Reference†
101	76.02(0.21)ǂ	76.96(0.22)	77.43(0.24)†	77.28(0.24)	1.71ǂ	0.62	Reference†	-
76	76.60(0.21)ǂ	77.51(0.22)	77.83(0.26)†	77.47(0.26)	1.43ǂ	0.50	Reference†	-
60	77.16(0.19)ǂ	77.71(0.21)	77.95(0.25)†	77.45(0.26)	1.11ǂ	0.35	Reference†	-
40	77.39(0.18)ǂ	77.71(0.24)	77.93(0.26)†	-	0.64ǂ	0.29	Reference†	-
32	77.71(0.23)ǂ	78.17(0.23)†	78.15(0.23)	-	0.70ǂ	Reference†	-	-
21	77.86(0.23)ǂ	78.28(0.23)	78.43(0.25)†	-	0.58ǂ	0.20	Reference†	-
15	78.54(0.23)ǂ	78.58(0.24)†	-	-	0.05ǂ	Reference†	-	-
5	77.68(0.23)ǂ	79.02(0.31)†	-	-	0.99ǂ	Reference†	-	-
2	77.92(0.30)	-	-	-	-	-	-	-
Suicide plans (n_{train}=4076, *ntree*=500)								
92	79.59(0.23)	80.61(0.24)	82.54(0.39)ǂ	85.52(0.34)†	2.84**	2.14*	1.21ǂ	Reference†
52	80.56(0.24)	82.55(0.27)	84.74(0.26)		-	-	-	-
Suicide Attempts (n_{train}=8153, *ntree*=500)								
105	77.48(0.47)ǂ	78.91(0.62)	80.04(0.76)	81.30(0.80)†	1.07ǂ	0.78	0.49	Reference†
76	77.84(0.50)ǂ	78.87(0.69)	80.13(0.70)	80.73(0.80)†	0.73ǂ	0.57	0.22	Reference†
52	78.09(0.50)ǂ	79.65(0.79)	80.74(0.70)†	-	0.84ǂ	0.41	Reference†	-
45	78.78(0.58)ǂ	79.52(0.73)	80.65(0.66)†	-	0.71ǂ	0.56	Reference†	-
36	79.48(0.67)ǂ	80.35(0.73)	80.78(0.76)†	-	0.42ǂ	0.16	Reference†	-
26	80.11(0.65)ǂ	80.78(0.72)	81.13(0.81)†	-	0.32ǂ	0.16	Reference†	-
15	82.17(0.68)†ǂ	82.13(0.66)	-	-	Reference†ǂ	-	-	-
4	83.78(0.60)†ǂ	-	-	-	Reference†ǂ	-	-	-
2	79.52(0.65)	-	-	-	-	-	-	-

[1]Including a subset of cross-sectional predictors; [2]Averaged over 50 iterations; [3]Corrected resampled t-tests (two-tail, df=49, α=0.05) comparing hyper-parameter tuned predictive performance of best† conditional random forest reduced model and all antecedent[4] reduced models; †Model with the highest predictive accuracy; [4]Model for which the values of *mtry* and/or *ntree* are lower than the corresponding values of the best reduced model; ; ǂOptimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown in a forest; p= number of variables in the model; n_{train}=number of observations used for training the classifier;*P< 0.05;**P< 0.01;***P< 0.001.

Table 7.5. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicidal ideation in Canadian general population

Full[3] Model (p=290, $mtry$=68)			1st Reduced[4] Model (p=157, $mtry$=24)			2nd Reduced[4] Model (p=115, $mtry$=10)			3rd Reduced[4] Model (p=81, $mtry$=9)		
Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)
284	MHS	2.35(0.39)	284	MHS	1.75(0.27)	284	MHS	1.17(0.14)	284	MHS	1.30(0.19)
121	MED	1.56(0.59)	121	MED	1.13(0.40)	121	MED	0.78(0.24)	121	MED	0.82(0.25)
229	MHE	0.90(0.38)	229	MHE	0.67(0.25)	229	MHE	0.50(0.15)	229	MHE	0.47(0.15)
212	DEP	0.72(0.23)	212	DEP	0.41(0.13)	46	GAD	0.29(0.18)	46	GAD	0.28(0.18)
46	GAD	0.45(0.39)	48	GAD	0.37(0.15)	48	GAD	0.28(0.09)	212	DEP	0.25(0.07)
48	GAD	0.43(0.22)	46	GAD	0.36(0.26)	212	DEP	0.25(0.08)	48	GAD	0.24(0.10)
64	GAD	0.35(0.24)	138	PNC	0.30(0.12)	138	PNC	0.25(0.08)	64	GAD	0.22(0.11)
221	DEP	0.35(0.19)	64	GAD	0.29(0.17)	64	GAD	0.24(0.11)	138	PNC	0.21(0.07)
138	PNC	0.31(0.16)	221	DEP	0.22(0.11)	43	GAD	0.15(0.11)	159	SPS	0.15(0.10)
211	DEP	0.27(0.19)	211	DEP	0.18(0.11)	221	DEP	0.15(0.07)	43	GAD	0.13(0.11)

4th Reduced[4] Model (p=60, $mtry$=7)			5th Reduced[4] Model (p=50, $mtry$=7)			6th Reduced[4] Model (p=46, $mtry$=6)			7th Reduced[4] Model (p=36, $mtry$=6)		
Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)
284	MHS	1.30(0.16)	284	MHS	1.41(0.18)	284	MHS	1.34(0.16)	284	MHS	1.45(0.19)
121	MED	0.81(0.24)	121	MED	0.86(0.25)	121	MED	0.83(0.25)	121	MED	0.75(0.27)
229	MHE	0.44(0.17)	229	MHE	0.44(0.16)	229	MHE	0.42(0.14)	229	MHE	0.36(0.15)
46	GAD	0.27(0.17)	46	GAD	0.30(0.21)	46	GAD	0.28(0.18)	46	GAD	0.28(0.20)
48	GAD	0.24(0.09)	48	GAD	0.25(0.10)	48	GAD	0.24(0.09)	212	DEP	0.25(0.08)
64	GAD	0.23(0.11)	212	DEP	0.22(0.07)	64	GAD	0.22(0.11)	64	GAD	0.22(0.12)
212	DEP	0.22(0.07)	64	GAD	0.22(0.11)	212	DEP	0.21(0.07)	48	GAD	0.21(0.09)
138	PNC	0.18(0.07)	138	PNC	0.17(0.08)	138	PNC	0.16(0.06)	159	SPS	0.17(0.11)
159	SPS	0.16(0.10)	159	SPS	0.16(0.11)	159	SPS	0.13(0.08)	28	AUD	0.15(0.09)
43	GAD	0.13(0.11)	28	AUD	0.13(0.09)	28	AUD	0.12(0.08)	163	SPS	0.12(0.09)

8th Reduced[4] Model (p=27, $mtry$=5)			9th Reduced[4] Model (p=22, $mtry$=4)			10th Reduced[4,6] Model (p=14, $mtry$=3)		
Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)	Variable: Index	Variable: Block	Mean decrease in predictive accuracy[5] %(SD)
284	MHS	1.50(0.22)	284	MHS	1.61(0.20)	284	MHS	1.97(0.25)
121	MED	0.82(0.29)	121	MED	0.83(0.28)	121	MED	1.31(0.33)
229	MHE	0.36(0.15)	229	MHE	0.37(0.15)	212	DEP	0.65(0.14)
46	GAD	0.30(0.21)	46	GAD	0.32(0.23)	64	GAD	0.54(0.20)
212	DEP	0.30(0.10)	64	GAD	0.32(0.15)	229	MHE	0.53(0.16)
48	GAD	0.25(0.11)	212	DEP	0.31(0.10)	211	DEP	0.43(0.15)
64	GAD	0.23(0.13)	48	GAD	0.27(0.11)	46	GAD	0.35(0.24)
28	AUD	0.17(0.10)	28	AUD	0.21(0.12)	28	AUD	0.34(0.16)
159	SPS	0.15(0.12)	159	SPS	0.18(0.13)	48	GAD	0.27(0.13)
221	DEP	0.15(0.07)	163	SPS	0.17(0.12)	268	MHS	0.20(0.12)

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy; [2]n_{train}=6260, *ntree*=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier; Abbreviations: AUD=Alcohol Use, Abuse, and Dependence; DEP=Depression; GAD=Generalized Anxiety Disorder; MED=Medication Use; MHE=Mental Health Experiences; MHS=Mental Health Services; PNC=Perceived Need for Care; SPS=Social Provision Scale.

Table 7.6. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicide plans in Canadian general population.

Full[3] Model (p=294,$mtry$=136)			**1st Reduced[4] Model** (p=125,$mtry$=88)			**2nd Reduced[4] Model** (p=53,$mtry$=28)			**3rd Reduced[4,6] Model** (p=34,$mtry$=20)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**		**Index**	**Block**		**Index**	**Block**	
294	SUI	13.34(0.55)	294	SUI	16.51(0.71)	294	SUI	13.10(0.97)	294	SUI	12.99(1.08)
292	SUI	10.32(1.53)	292	SUI	15.06(2.69)	292	SUI	8.06(1.57)	292	SUI	7.72(1.52)
212	DEP	2.21(1.30)	212	DEP	1.73(1.43)	212	DEP	1.51(1.09)	212	DEP	1.38(1.02)
211	DEP	1.02(1.02)	211	DEP	0.64(1.17)	211	DEP	0.69(0.73)	211	DEP	0.58(0.78)
221	DEP	0.56(0.52)	221	DEP	0.26(0.30)	221	DEP	0.32(0.34)	181	DEP	0.24(0.39)
181	DEP	0.53(0.66)	181	DEP	0.25(0.46)	284	MHS	0.31(0.20)	284	MHS	0.23(0.15)
284	MHS	0.49(0.30)	284	MHS	0.22(0.17)	181	DEP	0.27(0.37)	121	MED	0.22(0.21)
121	MED	0.37(0.27)	121	MED	0.16(0.18)	121	MED	0.27(0.22)	221	DEP	0.19(0.19)
171	DEP	0.18(0.38)	184	DEP	0.10(0.49)	184	DEP	0.10(0.32)	184	DEP	0.07(0.25)
184	DEP	0.18(0.46)	291	SUI	0.07(0.05)	45	GAD	0.08(0.12)	229	MHE	0.05(0.07)

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy;[2]n_{train}=12518, *ntree*=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; p=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier; Abbreviations: DEP=Depression; GAD=Generalized Anxiety Disorder; MED=Medication Use; MHE=Mental Health Experiences; MHS=Mental Health Services; SUI=Suicide.

Table 7.7. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicide attempts in Canadian general population.

Full[3] Model (p=298,$mtry$=68)			**1st Reduced[4,6] Model** (p=99,$mtry$=40)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**	
229	MHE	6.44(1.73)	229	MHE	8.07(2.56)
121	MED	4.00(1.64)	121	MED	4.21(2.22)
294	SUI	2.97(0.49)	294	SUI	3.95(1.08)
212	DEP	2.18(1.06)	212	DEP	2.25(1.90)
292	SUI	2.06(0.88)	292	SUI	2.10(1.47)
221	DEP	1.72(0.99)	221	DEP	1.54(1.28)
211	DEP	0.96(0.72)	211	DEP	0.76(0.84)
181	DEP	0.72(0.60)	298	SUI	0.58(0.24)
298	SUI	0.62(0.18)	181	DEP	0.51(0.56)
208	DEP	0.54(0.46)	296	SUI	0.31(0.40)

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy;[2]n_{train}=12518, *ntree*=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; p=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier; Abbreviations: DEP=Depression; MED=Medication Use; MHE=Mental Health Experiences; SUI=Suicide.

Table 7.8. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicidal ideation in Canadian Armed Forces.

Full[3] Model (p=289,$mtry$=68)			1st Reduced[4] Model (p=141,$mtry$=22)			2nd Reduced[4] Model (p=101,$mtry$=10)			3rd Reduced[4] Model (p=76,$mtry$=8)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**		**Index**	**Block**		**Index**	**Block**	
48	GAD	2.08(0.75)	48	GAD	1.40(0.44)	48	GAD	0.94(0.31)	48	GAD	0.81(0.26)
229	MHE	1.35(0.71)	229	MHE	0.86(0.43)	229	MHE	0.57(0.26)	229	MHE	0.56(0.27)
284	MHS	0.74(0.40)	284	MHS	0.60(0.28)	284	MHS	0.46(0.20)	284	MHS	0.44(0.21)
221	DEP	0.55(0.35)	221	DEP	0.37(0.20)	41	GAD	0.28(0.22)	221	DEP	0.24(0.11)
211	DEP	0.47(0.27)	211	DEP	0.31(0.17)	221	DEP	0.27(0.11)	41	GAD	0.22(0.17)
212	DEP	0.40(0.19)	41	GAD	0.27(0.24)	211	DEP	0.22(0.10)	211	DEP	0.20(0.09)
41	GAD	0.40(0.35)	212	DEP	0.20(0.12)	212	DEP	0.21(0.08)	212	DEP	0.19(0.08)
43	GAD	0.22(0.20)	43	GAD	0.15(0.16)	43	GAD	0.18(0.15)	274	MHS	0.15(0.08)
274	MHS	0.14(0.09)	274	MHS	0.13(0.08)	274	MHS	0.15(0.07)	46	GAD	0.13(0.09)
42	GAD	0.13(0.17)	46	GAD	0.12(0.11)	46	GAD	0.13(0.09)	43	GAD	0.13(0.11)

4th Reduced[4] Model (p=60,$mtry$=7)			5th Reduced[4] Model (p=40,$mtry$=6)			6th Reduced[4] Model (p=32,$mtry$=5)			7th Reduced[4] Model (p=21,$mtry$=4)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**		**Index**	**Block**		**Index**	**Block**	
48	GAD	0.79(0.27)	48	GAD	0.82(0.30)	48	GAD	0.71(0.25)	229	MHE	0.83(0.38)
229	MHE	0.56(0.28)	229	MHE	0.60(0.32)	229	MHE	0.67(0.32)	284	MHS	0.66(0.25)
284	MHS	0.50(0.21)	284	MHS	0.52(0.22)	284	MHS	0.51(0.20)	48	GAD	0.53(0.14)
221	DEP	0.24(0.11)	221	DEP	0.27(0.14)	221	DEP	0.25(0.12)	221	DEP	0.26(0.13)
41	GAD	0.20(0.17)	211	DEP	0.21(0.12)	211	DEP	0.20(0.10)	211	DEP	0.23(0.22)
211	DEP	0.20(0.11)	212	DEP	0.20(0.08)	212	DEP	0.19(0.09)	32	AUD	0.20(0.09)
212	DEP	0.18(0.07)	41	GAD	0.17(0.16)	41	GAD	0.16(0.15)	212	DEP	0.18(0.16)
274	MHS	0.13(0.06)	32	AUD	0.14(0.16)	32	AUD	0.15(0.17)	41	GAD	0.15(0.05)
46	GAD	0.13(0.08)	274	MHS	0.11(0.07)	274	MHS	0.10(0.05)	274	MHS	0.11(0.06)
43	GAD	0.11(0.11)	46	GAD	0.10(0.08)	46	GAD	0.10(0.08)	46	GAD	0.07(0.06)

8th Reduced[4] Model (p=15,$mtry$=4)			9th Reduced[4,6] Model (p=5,$mtry$=2)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**	
229	MHE	1.17(0.56)	229	MHE	3.46(1.11)
48	GAD	0.70(0.30)	48	GAD	3.09(0.87)
284	MHS	0.65(0.25)	284	MHS	1.58(0.54)
221	DEP	0.28(0.18)	221	DEP	0.80(0.46)
211	DEP	0.26(0.15)	32	AUD	0.26(0.36)
212	DEP	0.22(0.12)	-	-	-
32	AUD	0.17(0.21)	-	-	-
41	GAD	0.14(0.15)	-	-	-
46	GAD	0.07(0.07)	-	-	-
43	GAD	0.05(0.09)	-	-	-

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy;[2]n_{train}=4077, *ntree*=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; p=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier; Abbreviations: AUD=Alcohol Use, Abuse, and Dependence; DEP=Depression; GAD=Generalized Anxiety Disorder; MED=Medication Use; MHE=Mental Health Experiences; MHS=Mental Health Services.

Table 7.9. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicide plans in Canadian Armed Forces.

Full[3] Model (p=293,$mtry$=136)			**1st Reduced[4,6] Model** (p=92,$mtry$=36)		
Variable		**Mean decrease in predictive accuracy[5]** %(SD)	**Variable**		**Mean decrease in predictive accuracy[5]** %(SD)
Index	**Block**		**Index**	**Block**	
294	SUI	5.91(1.46)	294	SUI	5.03(1.19)
221	DEP	4.83(2.87)	221	DEP	3.92(2.27)
292	SUI	4.31(2.57)	292	SUI	3.74(2.04)
48	GAD	3.47(2.99)	48	GAD	2.53(1.95)
229	MHE	2.98(2.62)	212	DEP	2.28(1.64)
212	DEP	2.55(1.92)	229	MHE	2.15(1.78)
211	DEP	1.44(1.72)	211	DEP	1.30(1.37)
121	MED	0.84(0.98)	121	MED	0.74(0.76)
206	DEP	0.42(0.53)	206	DEP	0.39(0.46)
185	DEP	0.26(0.44)	185	DEP	0.26(0.38)

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy; [2]n_{train}=4076, *ntree*=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier; Abbreviations: DEP=Depression; GAD=Generalized Anxiety Disorder; MED=Medication Use; MHE=Mental Health Experiences; SUI=Suicide.

Table 7.10. Permutation accuracy importance scores[1] for predictors derived from conditional random forests models[2] of suicide attempts in Canadian Armed Forces.

Full[3] Model (p=297, $mtry$=17)			**1st Reduced[4] Model** (p=105, $mtry$=10)			**2nd Reduced[4] Model** (p=76, $mtry$=8)			**3rd Reduced[4] Model** (p=52, $mtry$=7)		
Variable		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)
Index	**Class**		**Index**	**Class**		**Index**	**Class**		**Index**	**Class**	
229	MHE	1.84(0.45)	229	MHE	2.65(0.85)	229	MHE	2.94(1.15)	229	MHE	3.40(1.26)
48	GAD	1.58(0.58)	48	GAD	2.61(0.99)	48	GAD	2.87(1.14)	48	GAD	3.40(1.39)
121	MED	1.48(0.46)	292	SUI	1.83(0.52)	292	SUI	1.93(0.59)	292	SUI	2.19(0.69)
292	SUI	1.40(0.40)	121	MED	1.79(0.82)	121	MED	1.91(0.84)	121	MED	2.09(1.05)
43	GAD	0.89(0.55)	43	GAD	1.22(0.85)	43	GAD	1.36(0.93)	43	GAD	1.43(1.08)
211	DEP	0.86(0.37)	211	DEP	0.99(0.59)	211	DEP	1.01(0.61)	211	DEP	1.07(0.73)
221	DEP	0.68(0.32)	221	DEP	0.94(0.55)	221	DEP	0.93(0.52)	221	DEP	0.97(0.63)
75	DEP	0.58(0.36)	75	DEP	0.72(0.46)	75	DEP	0.72(0.51)	75	DEP	0.75(0.51)
212	DEP	0.57(0.31)	82	GAD	0.61(0.54)	82	GAD	0.64(0.59)	82	GAD	0.63(0.57)
193	DEP	0.56(0.39)	96	GAD	0.55(0.40)	296	SUI	0.55(0.28)	193	DEP	0.50(0.47)

4th Reduced[4] Model (p=45, $mtry$=6)			**5th Reduced[4] Model** (p=36, $mtry$=6)			**6th Reduced[4] Model** (p=26, $mtry$=5)			**7th Reduced[4] Model** (p=15, $mtry$=4)		
Variable		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)	**Variable**		**Mean decrease in accuracy** %(SD)
Index	**Class**		**Index**	**Class**		**Index**	**Class**		**Index**	**Class**	
229	MHE	3.34(1.42)	229	MHE	3.92(1.72)	229	MHE	4.41(1.89)	48	GAD	5.72(2.69)
48	GAD	3.26(1.36)	48	GAD	3.85(1.69)	48	GAD	4.40(2.03)	229	MHE	5.65(2.36)
292	SUI	2.22(0.71)	292	SUI	2.36(0.79)	121	MED	2.70(1.60)	292	SUI	3.35(1.08)
121	MED	2.17(1.03)	121	MED	2.35(1.21)	292	SUI	2.63(0.88)	121	MED	3.27(1.80)
43	GAD	1.40(1.00)	43	GAD	1.52(1.22)	43	GAD	1.76(1.39)	43	GAD	1.92(1.49)
211	DEP	1.03(0.71)	211	DEP	1.13(0.89)	211	DEP	1.15(0.84)	211	DEP	0.38(1.11)
221	DEP	1.00(0.62)	221	DEP	1.04(0.72)	221	DEP	1.11(0.75)	221	DEP	0.33(0.79)
75	DEP	0.70(0.49)	75	DEP	0.76(0.62)	75	DEP	0.82(0.65)	75	DEP	0.93(0.65)
82	GAD	0.60(0.58)	82	GAD	0.59(0.62)	82	GAD	0.68(0.77)	82	GAD	0.72(0.77)
71	GAD	0.53(0.58)	40	GAD	0.52(0.64)	193	DEP	0.52(0.58)	40	GAD	0.63(0.77)

8th Reduced[4,6] Model (p=4, $mtry$=2)		
Variable		**Mean decrease in accuracy** %(SD)
Index	**Class**	
48	GAD	11.52(4.66)
229	MHE	10.72(4.25)
121	MED	6.63(3.76)
43	GAD	3.60(3.08)
-	-	-
-	-	-
-	-	-
-	-	-
-	-	-
-	-	-

[1]Top 10 ranked variables according to their mean decrease in predictive accuracy;[2]n_{train}=8153, $ntree$=500; [3]Model including all predictors; [4]Model including a subset of predictors; [5]Averaged over 50 iterations; [6]Best reduced model is the reduced model with the highest predictive accuracy; p=number of variables in a classifier; $mtry$=number of randomly chosen variables at each split; $ntree$=number of trees grown; n_{train}=number of observations used to train the classifier; Abbreviations: DEP=Depression; GAD=Generalized Anxiety Disorder; MED=Medication Use; MHE=Mental Health Experiences; SUI=Suicide.

Table 7.11. Selection of final best nested conditional random forest model of suicidal ideation in Canadian general population.

Model[1]	***p***[6]	***mtry***	**Predictive Accuracy**[7] **%(SE)**	**Variable Index**
Best Reduced[2]	14	12	80.32(0.20)	284,121,212,64,229,211,46,28,48,268,163,159,45,43
Optimal Reduced[3]		3	79.23(0.18)	
1st Nested	2	1	78.29(0.15)	284,121
2nd Nested	3	2	77.04(0.32)	284,121,212
3rd Nested	4	2	76.22(0.25)	284,121,212,64
4th Nested	5	2	78.26(0.14)	284,121,212,64,229
5th Nested		4	79.89(0.23)	
6th Nested	6	2	76.64(0.22)	284,121,212,64,229,211
7th Nested		4	79.75(0.22)	
8th Nested	7	2	76.38(0.20)	284,121,212,64,229,211,46
9th Nested		4	79.56(0.22)	
10th Nested	8	3	78.79(0.17)	284,121,212,64,229,211,46,28
11th Nested		6	80.06(0.21)	
12th Nested	9	3	78.86(0.18)	284,121,212,64,229,211,46,28,48
Best Nested[4]		6	80.55(0.21)	

[1]n_{train}=6260,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]Best nested[5] model is the nested model with the highest predictive accuracy; [5]All nested models are models within the best reduced model; [6]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [7]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.12. Selection of final best nested conditional random forest model of suicide plans in Canadian general population.

Model[1]	***p***[6]	***mtry***	**Predictive Accuracy**[7] **%(SE)**	**Variable Index**
Best Reduced[2] Optimal Reduced[3]	34	20	94.64(0.14)	294,292,212,211,181,284,121,221,184,229,45,171,291,48,193,202,200,43,183,288,82,286,170,266,123,108,208,274,219,72,186,185,44,268
1st Nested	2	1	93.25(0.31)	294,292
2nd Nested	3	2	92.38(0.58)	294,292,212
3rd Nested	4	2	81.25(0.40)	294,292,212,211
4th Nested	5	2	81.23(0.14)	294,292,212,211,181
5th Nested		4	94.25(0.14)	
6th Nested	6	2	82.36(0.24)	294,292,212,211,181,284
Best Nested[4]		4	94.80(0.24)	

[1]n_{train}=12518,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]Best nested[5] model is the nested model with the highest predictive accuracy; [5]All nested models are models within the best reduced model; [6]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [7]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.13. Selection of final best nested conditional random forest model of suicide attempts in Canadian general population.

Model[1]	***p***[6]	***mtry***	**Predictive Accuracy**[7] **%(SE)**	**Variable Index**
Best Reduced[2] Optimal Reduced[3]	99	40	85.23(0.28)	229,121,294,212,292,221,211,298,181,296,208,284,209,193,123,200,170,171,48,184,185,186,204,268,183,201,45,63,202,230,111,205,206,3,207,14,47,210,165,231,138,247,198,43,291,64,96,118,122,70,191,82,16,49,7,40,113,219,50,75,241,31,38,285,68,71,8,69,109,17,26,266,30,33,46,53,54,79,149,159,164,172,216,222,235,227,102,77,74,226,92,100,145,274,108,42,1,228,286
1st Nested	2	1	83.05(0.36)	229,121
2nd Nested	3	2	83.54(0.48)	229,121,294
3rd Nested	4	2	85.17(0.24)	229,121,294,212
4th Nested	5	2	85.67(0.27)	229,121,294,212,292
5th Nested		4	84.09(0.37)	
Best Nested[4]	6	2	86.46(0.19)	229,121,294,212,292,221

[1]n_{train}=12518,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]Best nested[5] model is the nested model with the highest predictive accuracy; [5]All nested models are models within the best reduced model; [6]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [7]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.14. Selection of final best nested conditional random forest model of suicidal ideation in Canadian Armed Forces.

Model[1]	***P***[5]	***mtry***	**Predictive Accuracy**[6] **%(SE)**	**Variable Index**
Best Reduced[2]	5	2	79.02(0.31)	229,48,284,221,32
Optimal Reduced[3]		4	77.68(0.23)	
1st Nested[4]	2	1	77.92(0.30)	229,48
2nd Nested[4]	3	2	78.04(0.30)	229,48,284
3rd Nested[4]	4	2	77.47(0.28)	229,48,284,221

[1]n_{train}=4077,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]All nested models are models within the best reduced model; [5]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [6]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.15. Selection of final best nested conditional random forest model of suicide plans in Canadian Armed Forces.

Model[1]	***p***[6]	***mtry***	**Predictive Accuracy**[7] **%(SE)**	**Variable Index**
Best Reduced[2] Optimal Reduced[3]	92	72	85.52(0.34)	294,221,292,48,212,229,211,121,206,185,43,52,41,184,171,183,47,201,202,198,170,74,230,208,204,123,205,200,186,96,231,284,181,113,77,88,142,72,60,193,40,82,71,75,110,247,25,216,42,73,253,148,289,45,34,76,236,122,249,222,237,68,141,44,31,138,163,248,261,209,33,38,147,7,28,78,115,150,190,203,89,210,109,9,159,164,165,111,268,228,286,27
Best Nested[4]	2	1	90.67(0.34)	294,221

[1]n_{train}=4076,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]Best nested[5] model is the nested model with the highest predictive accuracy; [5]All nested models are models within the best reduced model; [6]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [7]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.16. Selection of final best nested conditional random forest model of suicide attempts in Canadian Armed Forces.

Model[1]	***p***[6]	***mtry***	**Predictive Accuracy**[7] **%(SE)**	**Variable Index**
Best Reduced[2] Optimal Reduced[3]	4	2	83.78(0.60)	48,229,121,43
1st Nested	2	1	79.52(0.65)	48,229
Best Nested[4]	3	2	84.31(0.51)	48,229,121

[1]n_{train}=8153,*ntree*=500; [2]Best reduced model is the reduced model with the highest predictive accuracy; [3]Optimal reduced model is the least performant reduced model for which the null hypothesis of no difference between its predictive accuracy and the accuracy of the best reduced model cannot be rejected; [4]Best nested[5] model is the nested model with the highest predictive accuracy; [5]All nested models are models within the best reduced model; [6]Predictors of the best reduced model are added, one at a time, according to their variable importance scores; [7]Averaged over 50 iterations; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Table 7.17a. Performance measures during external validation of final conditional random forests models[1] of suicidal behaviour in Canadian general population and Canadian Armed Forces.

Past 12-month outcome	***p***	***mtry***	**Sample**				**Accuracy** %(SE)	**Sensitivity** %(SE)	**Specificity** %(SE)	**Positive predictive value** %(SE)	**Negative predictive value** %(SE)	**AUC**
			Train[2]	***n***[3]	**Test**[4]	***n***[3]						
Suicidal ideation	9	6	CIV1	25038	CAF1	8154	81.98(0.43)	79.54(0.45)	82.09(0.42)	16.49(0.41)	98.90(0.12)	0.8812
	9	6	CIV2	12519	CIV2	12519	87.38(0.30)	64.10(0.43)	88.26(0.29)	17.04(0.34)	98.49(0.11)	0.8505
	5	4	CAF1	8154	CIV1	25038	92.95(0.16)	51.45(0.32)	94.55(0.14)	26.66(0.28)	98.06(0.09)	0.8391
	5	4	CAF2	4077	CAF2	4077	88.91(0.49)	64.94(0.75)	89.98(0.47)	22.42(0.65)	98.29(0.20)	0.8728
Suicide Plans	6	4	CIV1	25034	CAF1	8152	95.33(0.23)	58.27(0.55)	95.91(0.22)	18.41(0.43)	99.31(0.09)	0.9520
	6	4	CIV2	12517	CIV2	12517	97.79(0.13)	68.97(0.41)	98.12(0.12)	30.12(0.41)	99.63(0.05)	0.9793
	2	1	CAF1	8152	CIV1	25034	95.40(0.13)	58.24(0.31)	95.83(0.13)	13.84(0.22)	99.50(0.04)	0.8580
	2	1	CAF2	4076	CAF2	4076	93.08(0.40)	67.16(0.74)	93.52(0.39)	14.75(0.56)	99.42(0.12)	0.9072
Suicide attempts	6	4	CIV1	25036	CAF1	8153	94.30(0.26)	82.61(0.42)	94.33(0.26)	3.96(0.22)	99.95(0.25)	0.9786
	6	4	CIV2	12518	CIV2	12518	96.33(0.17)	65.38(0.43)	96.45(0.17)	7.14(0.23)	99.85(0.03)	0.9404
	3	2	CAF1	8153	CIV1	25036	91.78(0.17)	66.96(0.30)	91.89(0.17)	3.67(0.12)	99.83(0.03)	0.8334
	3	2	CAF2	4077	CAF2	4076	99.71(0.03)	0.00	100.00	-	99.71(0.03)	0.5000

[1]Nested model with the highest predictive accuracy; [2]Balanced; [3]Unweighted; [4]Unbalanced; Abbreviations: CIV1=entire CCHS-MH sample; CIV2=half of CCHS-MH sample; CAF1=entire CFMHS sample; CAF2=half of CFMHS sample; SE=standard error estimated by exact binomial test; AUC=Area under the ROC curve.

Table 7.17b. Classification accuracy during external validation of final conditional random forests models[2] of suicidal behaviour in Canadian general population and Canadian Armed Forces.

Past 12-month outcome	***p***	***mtry***	**Sample**				***n***[3]			
			Train[3]	***n***[3]	**Test**[4]	***n***[3]	**True positives**	**False negatives**	**True negatives**	**False positives**
Suicidal ideation	9	6	CIV1	25038	CAF1	8154	276	71	6409	1398
	9	6	CIV2	12519	CIV2	12519	291	163	10648	1417
	5	4	CAF1	8154	CIV1	25038	478	451	22794	1315
	5	4	CAF2	4077	CAF2	4077	113	61	3512	391
Suicide plans	6	4	CIV1	25034	CAF1	8152	74	53	7697	328
	6	4	CIV2	12517	CIV2	12517	100	45	12140	232
	2	1	CAF1	8152	CIV1	25034	166	119	23716	1033
	2	1	CAF2	4076	CAF2	4076	45	22	3749	260
Suicide attempts	6	4	CIV1	25036	CAF1	8153	19	4	7669	461
	6	4	CIV2	12518	CIV2	12518	34	18	12024	442
	3	2	CAF1	8153	CIV1	25036	77	38	22901	2020
	3	2	CAF2	4077	CAF2	4076	0	12	4064	0

[1]Nested model with the highest predictive accuracy; [2]Balanced; [3]Unweighted; [4]Unbalanced; Abbreviations: CIV1=entire CCHS-MH sample; CIV2=half of CCHS-MH sample; CAF1=entire CFMHS sample; CAF2=half of CFMHS sample.

Table 7.18. Population-specific common† cross-sectional predictors‡ of suicide behaviour.

Population	Variable			
	Index	**Block**	**Recall period**	**Description**
Canadian general population	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs
	212	DEP	Lifetime	Thought it would be better if dead[1]
	229	MHE	12 months	Received treatment for emotional or mental health problem
	211	DEP	Lifetime	Thought a lot about own death or death in general[1]
	48	GAD	Lifetime	Worried about own mental health
	268	MHS	12 months	Talked with a friend at least once a month about own problems with emotions, mental health or use of alcohol or drugs
	45	GAD	Lifetime	Worried about relationships with family
	43	GAD	Lifetime	Worried about love life
Canadian Armed Forces	229	MHE	12 months	Received treatment for emotional or mental health problem
	48	GAD	Lifetime	Worried about own mental health

†Common to all suicide behaviour outcomes; ‡Predictors derived from best reduced conditional random forests models; [1]Symptom experienced during a period of 2 weeks or longer, in which the respondent had reported as being sad, discouraged or being uninterested nearly every day; Abbreviations: MHS=Mental Health Services; MED=Medication Use; DEP=Depression; MHE=Mental Health Experiences; GAD=Generalized Anxiety Disorder.

Table 7.19. Common† cross-sectional predictors‡ of suicide behaviour in Canadian general population and Canadian Armed Forces

Past 12-month outcome	Variable			
	Index	**Block**	**Recall period**	**Description**
Suicidal ideation	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
	229	MHE	12 months	Received treatment for emotional or mental health problem
	48	GAD	Lifetime	Worried about own mental health
Suicide plans	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs
	212	DEP	Lifetime	Thought it would be better if dead[1]
	229	MHE	12 months	Received treatment for emotional or mental health problem
	211	DEP	Lifetime	Thought a lot about own death or death in general[1]
	48	GAD	Lifetime	Worried about own mental health
	268	MHS	12 months	Talked with a friend at least once a month about own problems with emotions, mental health or use of alcohol or drugs
	45	GAD	Lifetime	Worried about relationships with family
	43	GAD	Lifetime	Worried about love life
	221	DEP	12 months	Major depressive episode with significant interference on daily activities and responsibilities[3]
	292	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation during worst episode[2]
	294	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[2]
	181	DEP	Lifetime	Felt so sad that nothing could cheer up[1]
	184	DEP	Lifetime	Lost interest in almost all things[1]
	171	DEP	Lifetime	Had depressive-like symptoms[1]
	193	DEP	Lifetime	Felt tired/low in energy[1]
	202	DEP	Lifetime	Lost self-confidence[1]
	200	DEP	Lifetime	Had more trouble concentrating than usual[1]
	183	DEP	Lifetime	Felt hopeless about the future[1]
	82	GAD	Lifetime	Felt sad or depressed most of the time[4]
	286	CEX	Childhood	Experienced severe physical abuse in childhood[5]
	170	DEP	Lifetime	Experienced severe emotional distress[1]
	123	PNC	12 months	Received as much help as needed in 1 or 2 (but not all) of the following 3 domains: information, medication, or counseling
	208	DEP	Lifetime	Unable to cope with everyday responsibilities[1]
	72	GAD	Lifetime	Felt restless[4]
	186	DEP	Lifetime	Significant weight loss or gain or change in appetite[1]
	185	DEP	Lifetime	Felt like nothing was fun[1]
	44	GAD	Lifetime	Worried about relationships at school or work
Suicide attempts	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs
	229	MHE	12 months	Received treatment for emotional or mental health problem
	48	GAD	Lifetime	Worried about own mental health
	43	GAD	Lifetime	Worried about love life

†Common to each suicide behaviour outcome, regardless of population setting; †Predictors derived from best reduced conditional random forests models; [1]Symptom experienced during a period of 2 weeks or longer, in which the respondent had reported as being sad, discouraged or being uninterested nearly every day; [2]A period of 2 weeks of longer when feelings of being sad, discouraged or uninterested were most serious and frequent; [3]Normal routine, occupational (academic) functioning, or social activities or relationships; [4] Symptom experienced during a period lasting 6 months or longer of feeling worried most days;[5] Respondents aged 20 and over reporting as being physically attacked at least once or being slapped/hit/spanked at least 10 times, or being pushed/grabbed/shoved at least 10 times before age 16; Abbreviations: MHS=Mental Health Services; MED=Medication Use; DEP=Depression; MHE=Mental Health Experiences; GAD=Generalized Anxiety Disorder; SUI=Suicide; CEX=Adverse Childhood Experiences; PNC=Perceived Need for Care.

Table 7.20. Cross-sectional risk algorithms† for suicidal behaviour in Canadian general population

Past 12-month outcome	Variable			
	Index	Block	Recall period	Description
Suicidal ideation	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs
	212	DEP	Lifetime	Thought it would be better if dead[1]
	64	GAD	Lifetime	Often nervous or worried that could not think about something else
	229	MHE	12 months	Received treatment for emotional or mental health problem
	211	DEP	Lifetime	Thought a lot about own death or death in general[1]
	46	GAD	Lifetime	Worried about physical appearance
	28	AUD	Lifetime	Unable to cut down alcohol use despite making several attempts to quit
	48	GAD	Lifetime	Worried about own mental health
Suicide plans	292	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation during worst episode[2]
	294	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[2]
	212	DEP	Lifetime	Thought it would be better if dead[1]
	211	DEP	Lifetime	Thought a lot about own death or death in general[1]
	181	DEP	Lifetime	Felt so sad that nothing could cheer up[1]
	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
Suicide attempts	229	MHE	12 months	Received treatment for emotional or mental health problem
	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs
	294	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[3]
	212	DEP	Lifetime	Thought it would be better if dead[1]
	292	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation during worst episode[2]
	221	DEP	12 months	Major depressive episode with significant interference on daily activities and responsibilities[3]

†Predictors used in best nested (final) conditional random forest models; [1]Symptom experienced during a period of 2 weeks or longer, in which the respondent had reported as being sad, discouraged or being uninterested nearly every day; [2]A period of 2 weeks of longer when feelings of being sad, discouraged or uninterested were most serious and frequent; [3]Normal routine, occupational (academic) functioning, or social activities or relationships; Abbreviations: MHS=Mental Health Services; MED=Medication Use; DEP=Depression; MHE=Mental Health Experiences; GAD=Generalized Anxiety Disorder; AUD=Alcohol Use Disorders; SUI=Suicide.

Table 7.21. Cross-sectional risk algorithms† for suicidal behaviour in Canadian Armed Forces

Past 12-month outcome	Variable			
	Index	Block	Recall period	Description
Suicidal ideation	229	MHE	12 months	Received treatment for emotional or mental health problem
	48	GAD	Lifetime	Worried about own mental health
	284	MHS	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it
	221	DEP	12 months	Major depressive episode with significant interference on daily activities and responsibilities[1]
	32	AUD	Lifetime	Drinking alcohol frequently interfered with work or responsibilities at school, on a job, or at home
Suicide plans	294	SUI	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[2]
	221	DEP	12 months	Major depressive episode with significant interference on daily activities and responsibilities[1]
Suicide attempts	48	GAD	Lifetime	Worried about own mental health
	229	MHE	12 months	Received treatment for emotional or mental health problem
	121	MED	12 months	Received medication for emotions, mental health or use of alcohol or drugs

†Predictors used in best nested (final) conditional random forest models; [1]Normal routine, occupational (academic) functioning, or social activities or relationships;[2]A period of 2 weeks of longer when feelings of being sad, discouraged or uninterested were most serious and frequent; Abbreviations: MHS=Mental Health Services; MED=Medication Use; DEP=Depression; MHE=Mental Health Experiences; GAD=Generalized Anxiety Disorder; AUD=Alcohol Use Disorders; SUI=Suicide;

Table 7.22. Distribution of socio-demographic characteristics in Canadian general population and Canadian Armed Forces†

Socio-demographic characteristic	Canadian general population (n= 28,314,720)		Canadian Armed Forces (n=68,860)	
	Weighted Population	**Percentage (95%CI)**	**Weighted Population**	**Percentage (95%CI)**
Age, years				
Less than 20	2,294,200	8.10(7.74-8.47)	840	1.21(0.94-1.50)*
20 to 29	4,474,140	15.80(15.10-16.50)	21,860	31.74(30.64-32.85)*
30 to 39	4,447,940	15.71(15.04-16.38)	22,020	31.98(30.82-33.13)*
40 to 49	4,948,000	17.48(16.53-18.42)	17,460	25.36(24.42-26.29)*
50 years or more	12,150,420	42.91(42.28-43.54)	6,700	9.73(9.08-10.38)*
Male	13,954,740	49.28	59,540	86.47(85.66-87.27)*
Marital status				
Married	13,943,200	49.34(48.33-50.34)	31,000	45.05(43.89-46.20)*
Common Law	3,038,920	10.75(10.07-11.44)	13,880	20.17(19.71-21.16)*
Widowed	1,375,300	4.87(4.57-5.16)	120	0.17(0.09-0.26)*
Separated	824,980	2.92(2.48-3.36)	2,580	3.75(3.29-4.21)
Divorced	1,449,760	5.13(4.72-5.54)	2,440	3.55(3.12-3.98)*
Single	7,629,960	27.00(26.29-27.71)	18,820	27.35(26.32-28.37)
Ethnicity				
White	21,687,080	79.56(78.30-80.82)	61,900	90.10(89.31-90.89)*
Non-white	5,200,340	19.08(17.83-20.32)	4,220	6.14(5.52-6.77)*
Multiple origins	370,240	1.36(1.10-1.61)	2,600	3.78(3.30-4.27)*
Living alone	4,189,620	14.80(14.18-15.41)	16,060	23.32(22.28-24.37)*
Education				
Less than secondary	5,101,440	18.12(17.35-18.88)	2,720	3.96(3.48-4.44)*
Secondary graduation	4,441,300	15.77(15.02-16.52)	17,440	25.38(24.32-26.44)*
Some postsecondary	2,003,480	7.11(6.55-7.68)	6,120	8.91(8.24-9.57)*
Postsecondary graduation	16,612,660	59.00(57.92-60.07)	42,440	61.76(60.62-62.90)*
Employment status				
Unemployed	10,126,580	35.96(35.11-36.81)	0	0*
Full-time	14,849,800	52.74(51.78-53.70)	68,300	99.22(99.10-99.33)*
Part-time	3,136,120	11.14(10.55-11.73)	540	0.78(0.67-0.90)*
Difficult current income	3,302,160	11.76(11.00-12.51)	4,240	6.17(5.58-6.75)*
Personal income, $				
Less than 10,000	3,466,040	14.01(13.34-14.68)	200	0.29(0.16-0.42)*
10,000 to 19,999	3,941,680	15.93(15.18-16.68)	880	1.29(1.05-1.53)*
20,000 to 29,999	3,391,520	13.71(12.99-14.42)	940	1.38(1.11-1.65)*
30,000 to 39,999	3,106,280	12.56(11.78-13.33)	2,420	3.55(3.04-4.06)*
40,000 to 49,999	2,709,460	10.95(10.19-11.72)	4,200	6.16(5.51-6.80)*
50,000 to 59,999	2,066,920	8.35(7.71-9.00)	9,700	14.22(13.33-15.12)*
60,000 to 69,999	1,570,640	6.35(5.73-6.97)	17,500	25.66(24.65-26.67)*
70,000 to 79,999	1,230,780	4.98(4.47-5.48)	11,900	17.45(16.57-18.32)*
$80,000 to $89,999	828,540	3.35(2.96-3.74)	7,580	11.11(10.47-11.76)*
$90,000 to $99,999	580,960	2.35(1.98-2.71)	4,020	5.89(5.44-6.35)*
$100,000 or more	1,846,020	7.46(6.82-8.10)	8,880	13.02(12.43-13.61)*

CI=Confidence intervals; †Prevalence estimates were calculated using survey and bootstrap weights generated by Statistics Canada for each survey. The corresponding 95%CI for all prevalence estimates were obtained using bootstrap methods; *Proportions are significantly different if their 95%CI do not overlap.

Table 7.23. Prevalence of past 12-month suicidal behaviour outcomes across constructs of final risk algorithms in Canadian general population.

Past 12-month outcome	Construct: Recall period	Construct: Description	Conditional prevalence of past 12-month outcome† %(95%CI)
Suicidal ideation	12 months	None	3.34(3.00-3.69)
	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it	22.39(19.21-25.56)
	12 months	Received medication for emotions, mental health or use of alcohol or drugs	15.04(12.82-17.26)
	Lifetime	Thought it would be better if dead[1]	28.59(24.56-32.61)
	Lifetime	Often nervous or worried that could not think about something else	16.24(13.60-18.87)
	12 months	Received treatment for emotional or mental health problem	18.02(15.40-20.64)
	Lifetime	Thought a lot about own death or death in general[1]	21.72(18.63-24.81)
	Lifetime	Worried about physical appearance	11.70(9.53-13.86)
	Lifetime	Unable to cut down alcohol use despite making several attempts to quit	10.93(8.35-13.52)
	Lifetime	Worried about own mental health	19.46(16.40-22.53)
	Both	All of the above	76.00(34.52-100.00)[E]
Suicide plans	12 months	None	1.14(0.89-1.39)
	12 months	Past year suicidal ideation and history of lifetime suicidal ideation during worst episode[2]	20.07(13.73-26.42)
	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[2]	43.81(35.59-52.03)
	Lifetime	Thought it would be better if dead[1]	13.33(9.90-16.77)
	Lifetime	Thought a lot about own death or death in general[1]	10.36(7.73-12.99)
	Lifetime	Felt so sad that nothing could cheer up[1]	7.89(6.18-9.60)
	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it	9.88(7.31-12.46)
	Both	All of the above	53.37(43.55-63.19)
Suicide attempts	12 months	None	0.53(0.33-0.72)[E]
	12 months	Received treatment for emotional or mental health problem	4.51(2.52-6.50)[E]
	12 months	Received medication for emotions, mental health or use of alcohol or drugs	3.63(1.93-5.33)[E]
	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[3]	8.05(4.40-11.70)[E]
	Lifetime	Thought it would be better if dead[1]	7.01(3.76-10.25)[E]
	12 months	Past year suicidal ideation and history of lifetime suicidal ideation during worst episode[2]	21.98(12.76-31.19)[E]
	12 months	Major depressive episode with significant interference on daily activities and responsibilities[3]	7.82(4.01-11.63)[E]
	Both	All of the above	32.65(16.66-48.65)[E]

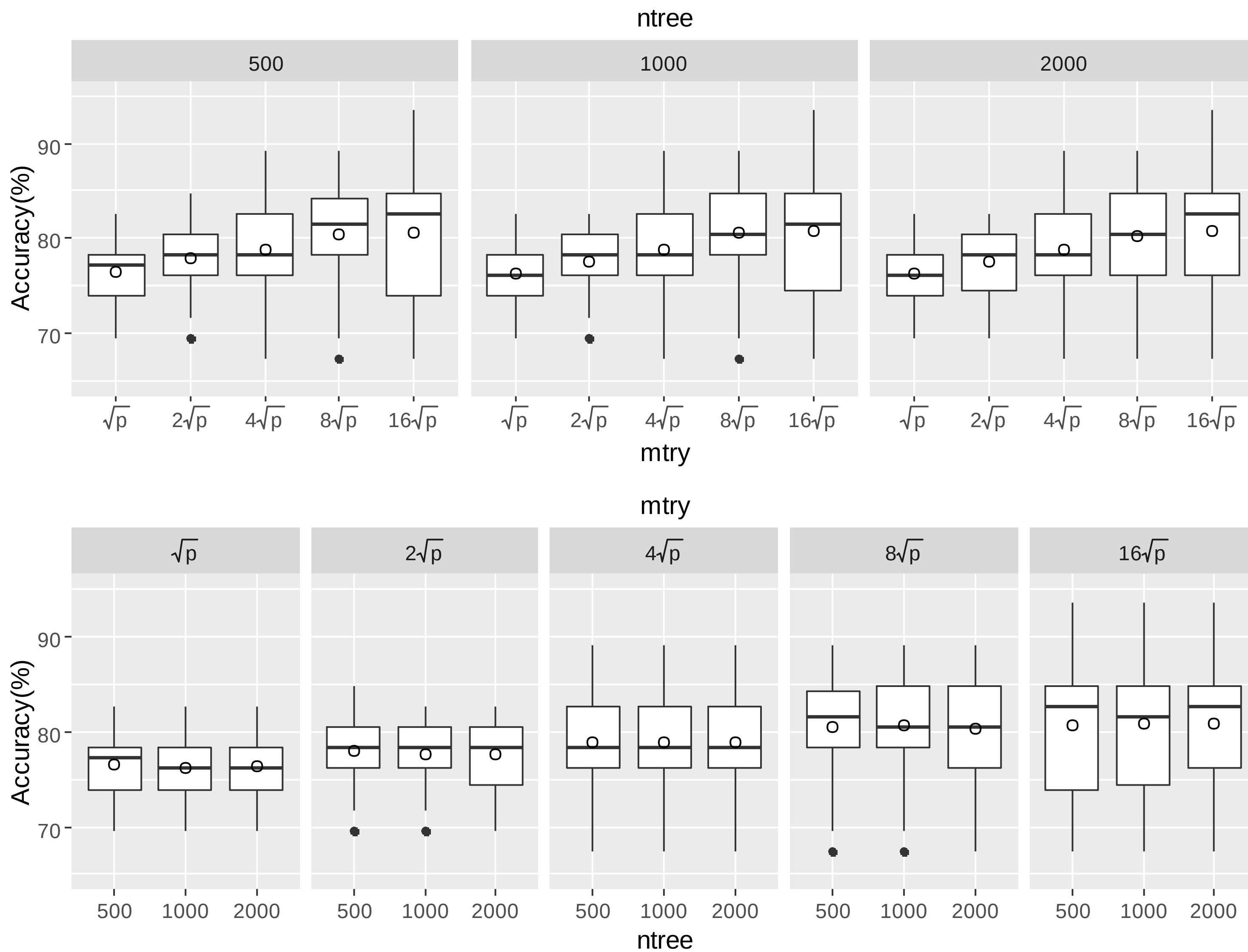
†Conditional on constructs of final risk algorithms; [1]Symptom experienced during a period of 2 weeks or longer, in which the respondent had reported as being sad, discouraged or being uninterested nearly every day; [2]A period of 2 weeks of longer when feelings of being sad, discouraged or uninterested were most serious and frequent; [3]Normal routine, occupational (academic) functioning, or social activities or relationships; [E]High sampling variability associated with the estimate.

Table 7.24. Prevalence of past 12-month suicidal behaviour outcomes across constructs of final risk algorithms in Canadian Armed Forces.

Past 12-month outcome	Construct: Recall period	Construct: Description	Conditional prevalence of past 12-month outcome† %(95%CI)
Suicidal ideation	12 months	None	4.36(3.83-4.89)
	12 months	Received treatment for emotional or mental health problem	17.29(14.75-19.83)
	Lifetime	Worried about own mental health	17.49(15.04-19.93)
	12 months	Felt that needed help for emotions, mental health or use of alcohol or drugs at least once, but did not receive it	17.20(13.93-20.46)
	12 months	Major depressive episode with significant interference on daily activities and responsibilities[1]	28.74(24.39-33.08)
	Lifetime	Drinking alcohol frequently interfered with work or responsibilities at school, on a job, or at home	9.55(7.42-11.67)
	Both	All of the above	36.36(16.02-56.70)[E]
Suicide plans	12 months	None	1.72(1.40-2.04)
	12 months	Past year suicidal ideation and history of lifetime suicidal ideation outside worst episode[2]	26.76(19.10-34.43)
	12 months	Major depressive episode with significant interference on daily activities and responsibilities[1]	15.38(11.97-18.80)
	Both	All of the above	80.00(48.62-100.00)[E]
Suicide attempts	12 month	None	0.35(0.18-0.52)[E]
	Lifetime	Worried about own mental health	2.00(1.06-2.95)[E]
	12 months	Received treatment for emotional or mental health problem	1.96(0.99-2.93)[E]
	12 months	Received medication for emotions, mental health or use of alcohol or drugs	2.34(1.12-3.56)[E]
	Both	All of the above	4.50(2.18-6.82)[E]

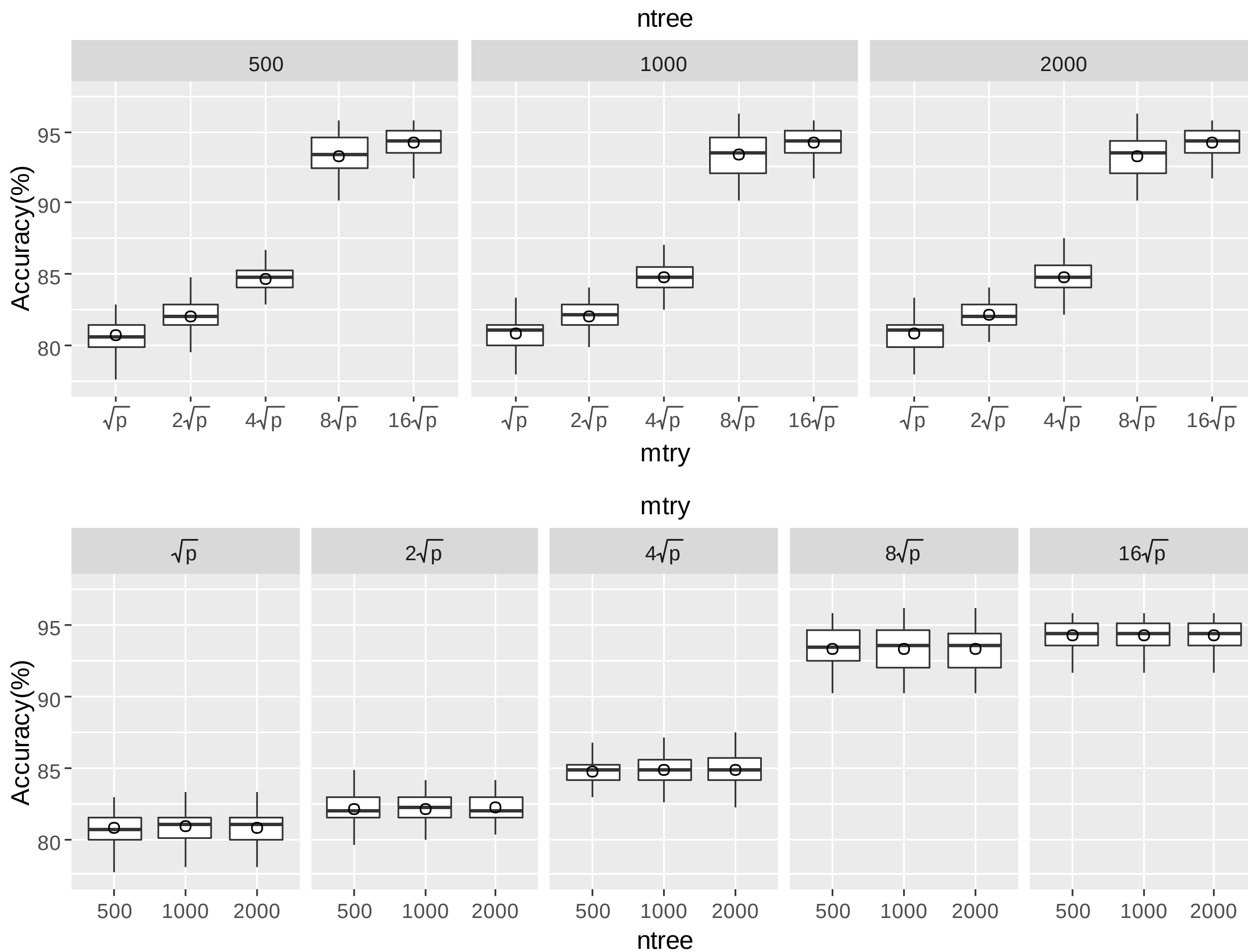
†Conditional on constructs of final risk algorithms; [1]Normal routine, occupational (academic) functioning, or social activities or relationships;[2]A period of 2 weeks of longer when feelings of being sad, discouraged or uninterested were most serious and frequent; ; [E]High sampling variability associated with the estimate.

Figure 7.1. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicidal ideation in Canadian general population.



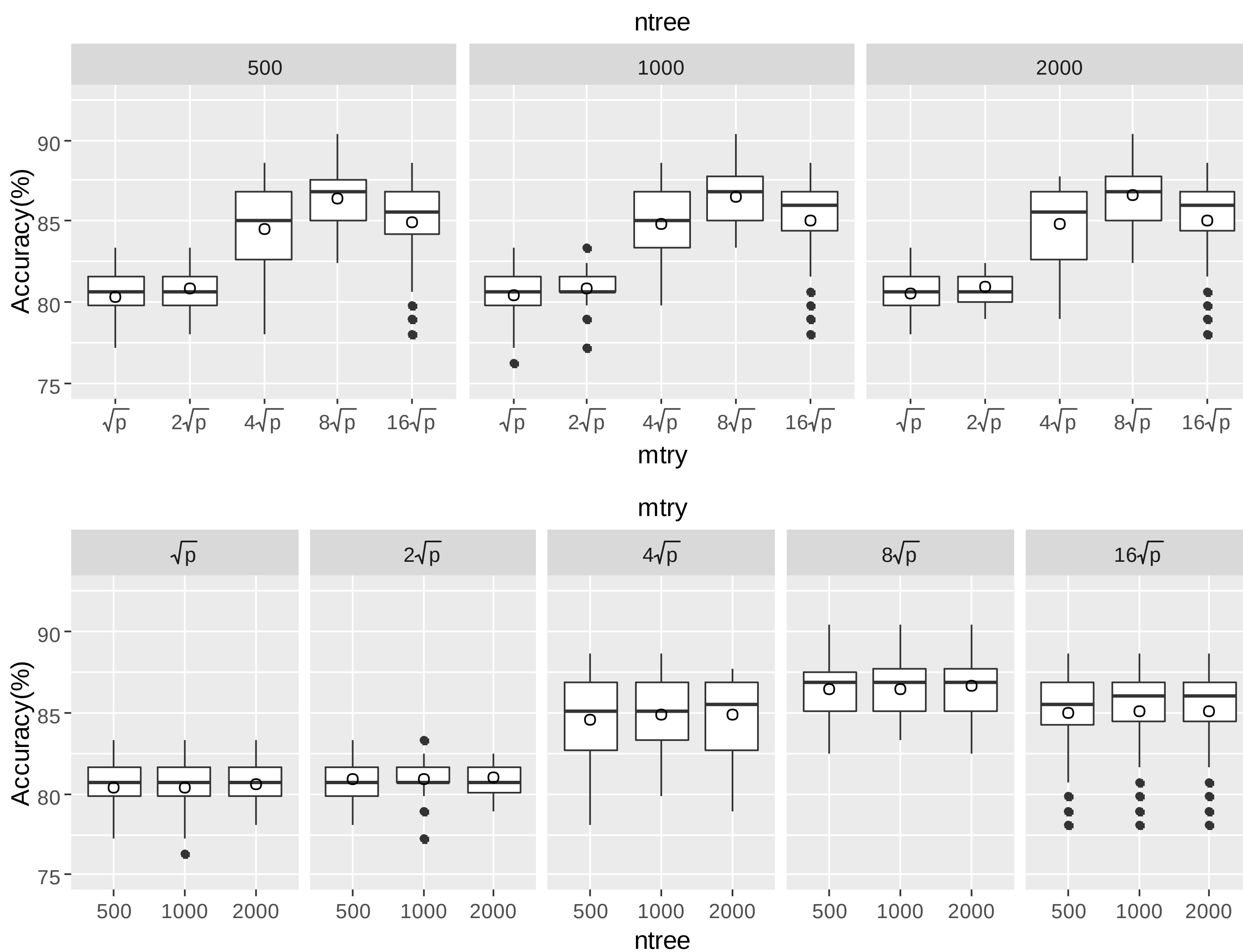
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.2. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicide plans in Canadian general population.



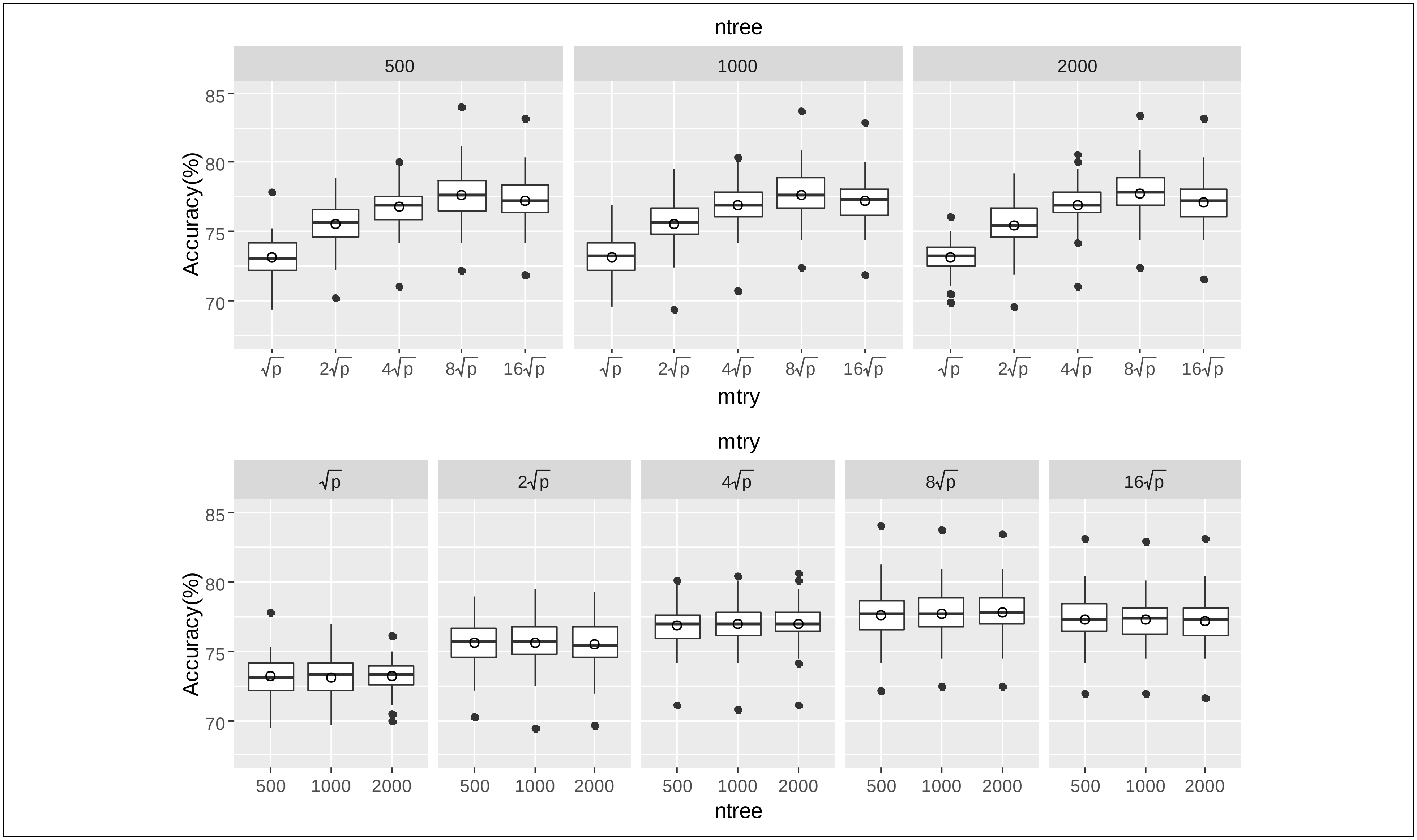
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.3. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicide attempts in Canadian general population.



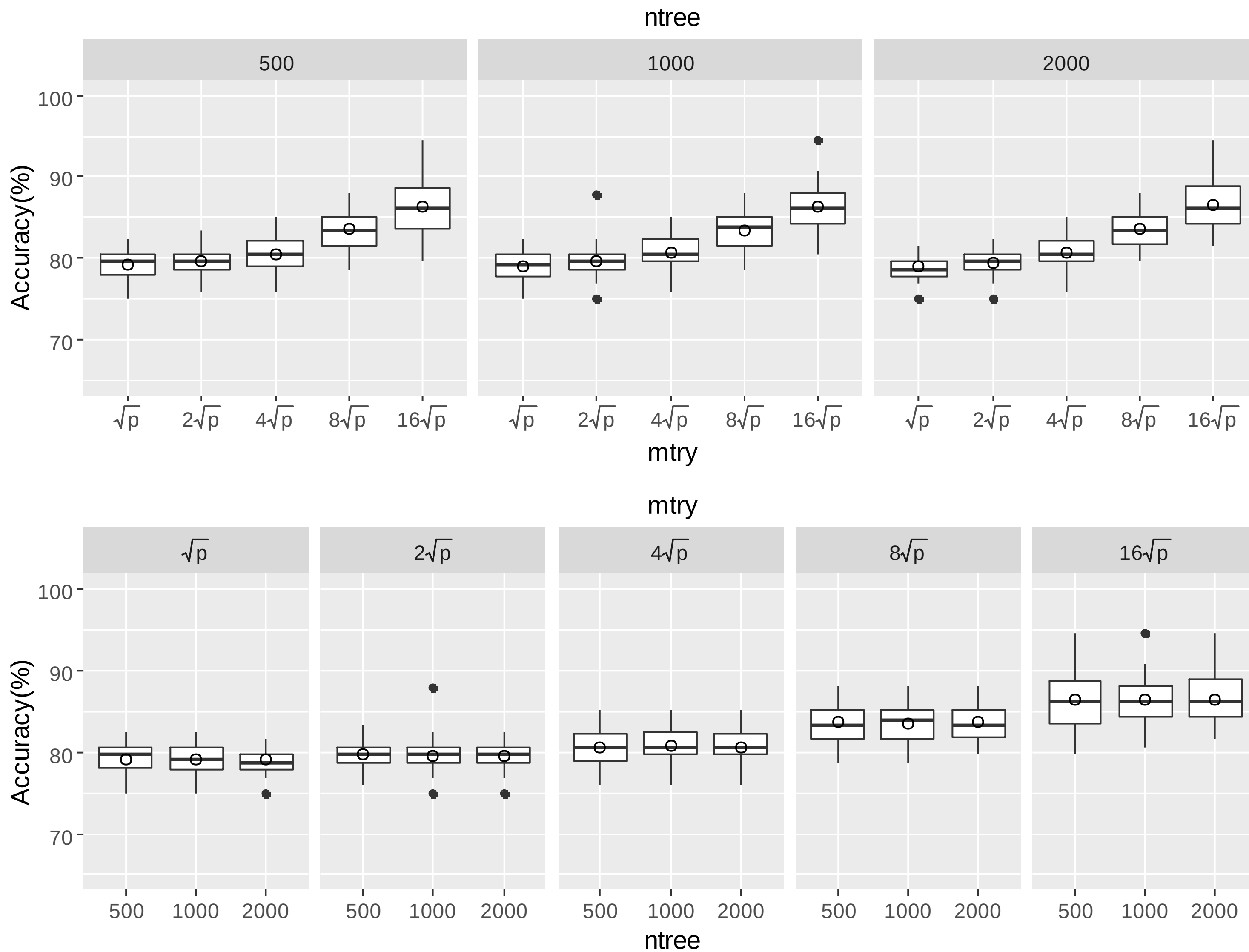
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.4. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicidal ideation in Canadian Armed Forces.



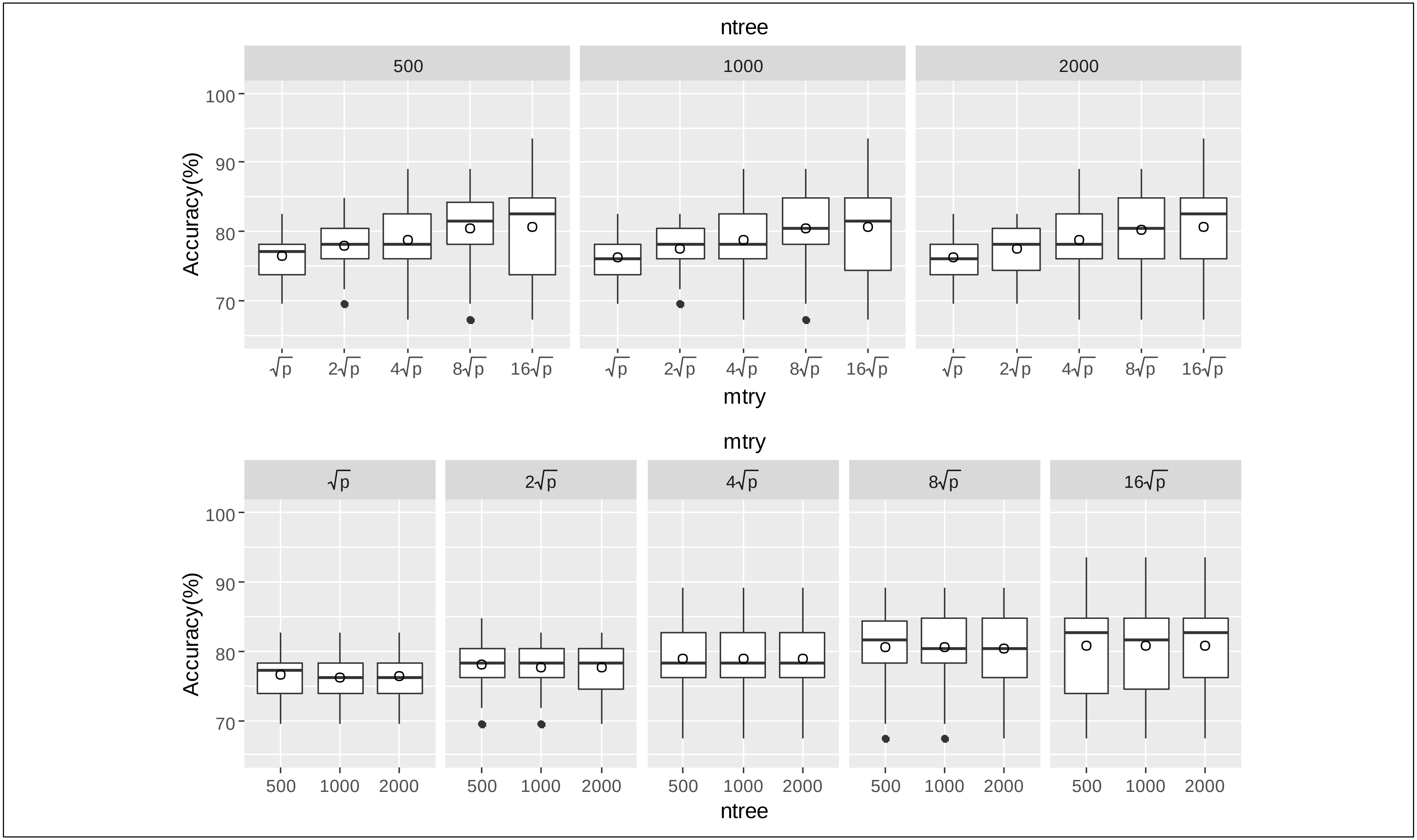
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.5. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicide plans in Canadian Armed Forces.



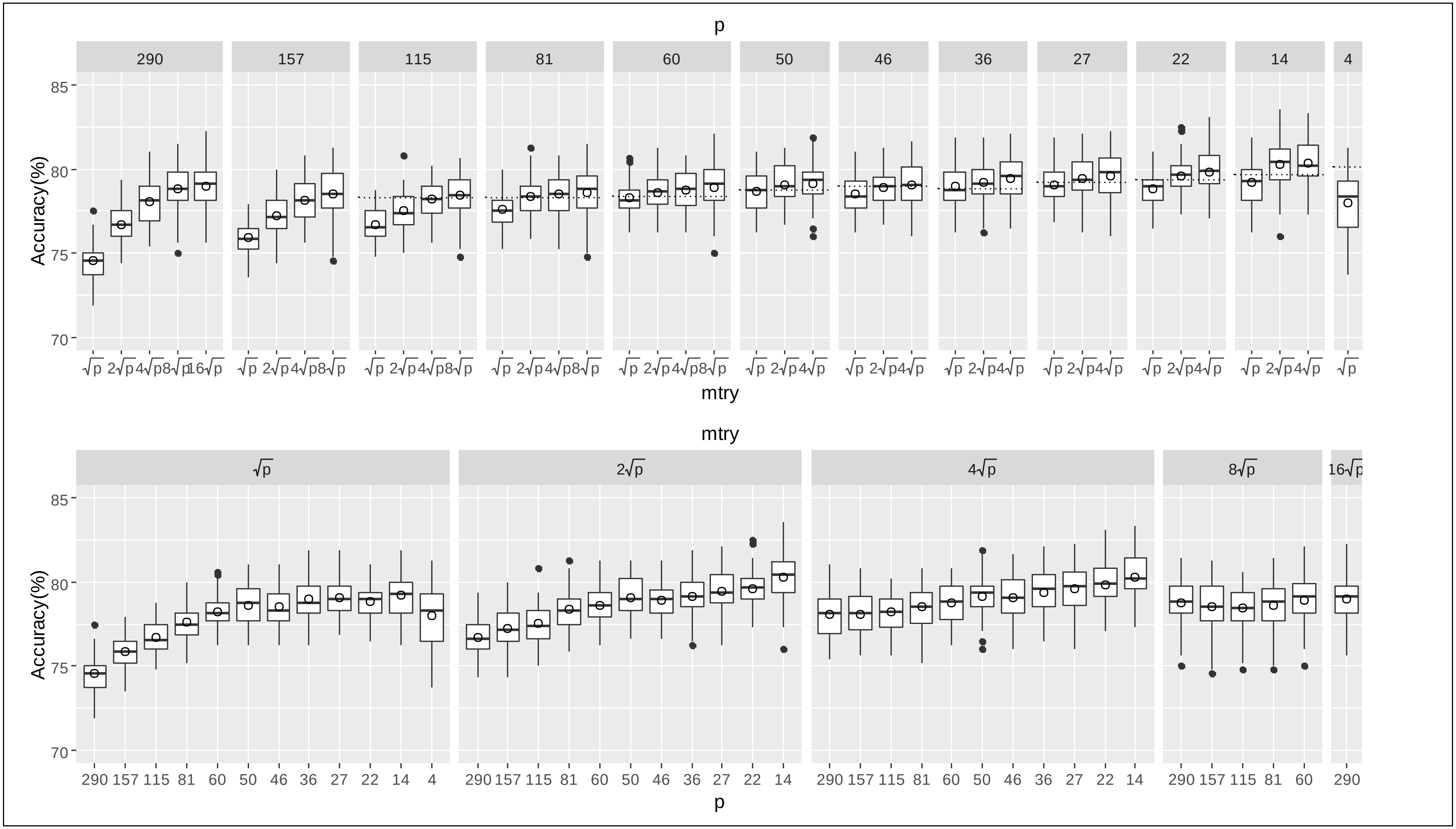
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.6. Box plots of predictive accuracies[1] for hyper-parameter tuned full[2] conditional random forests models of suicide attempts in Canadian Armed Forces.



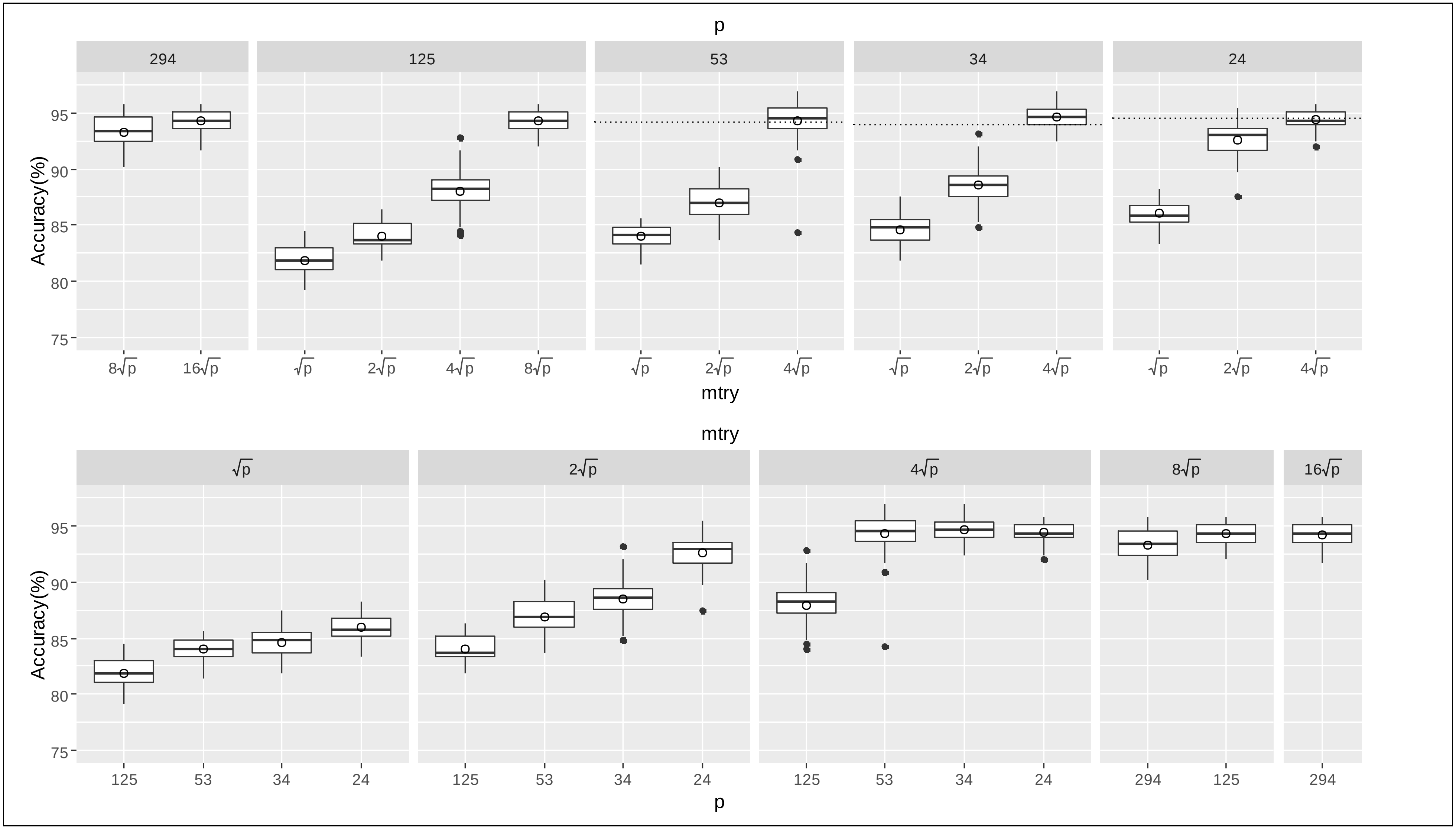
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model.

Figure 7.7. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicidal ideation in Canadian general population.



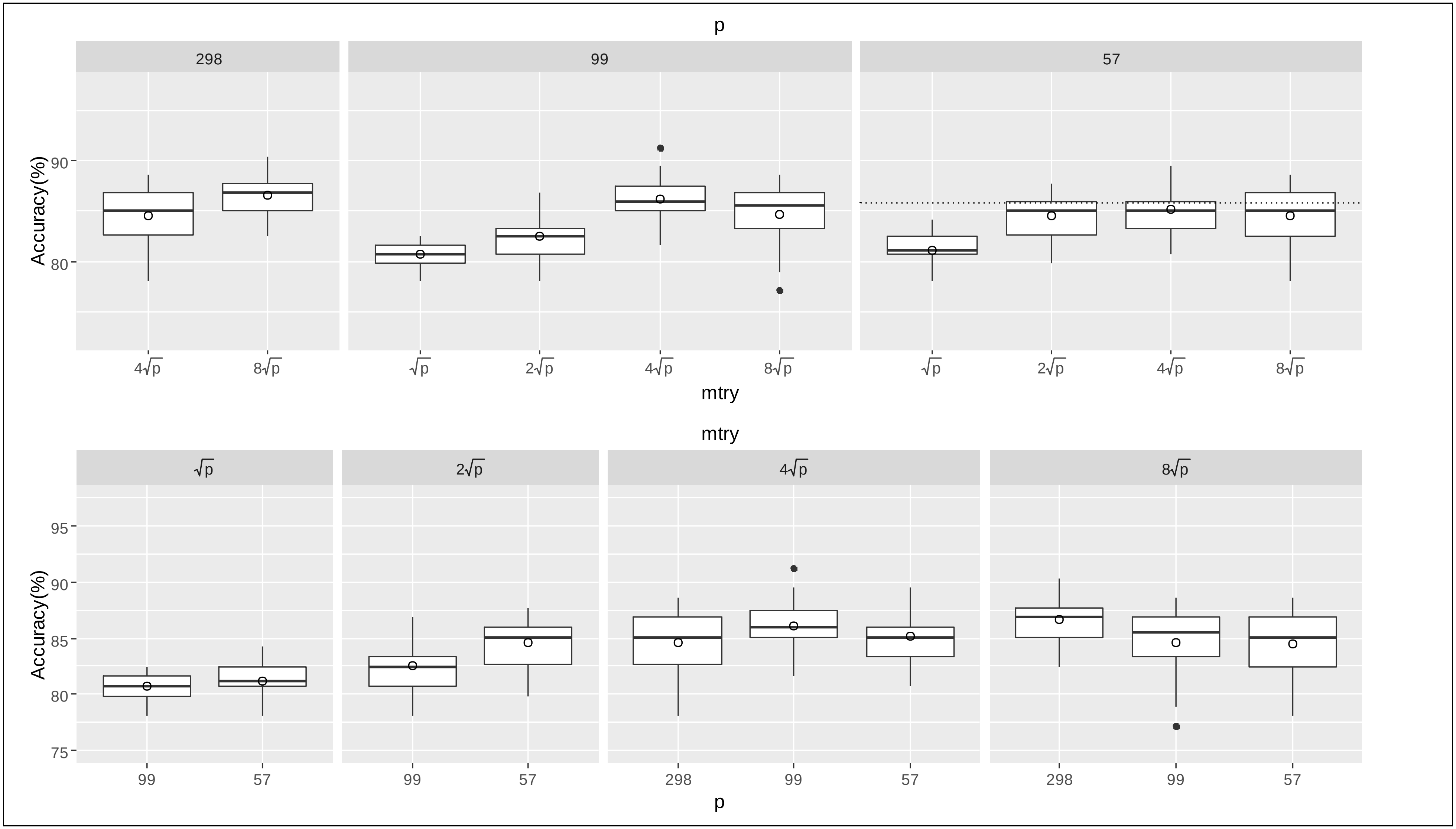
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model; Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model.

Figure 7.8. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicide plans in Canadian general population.



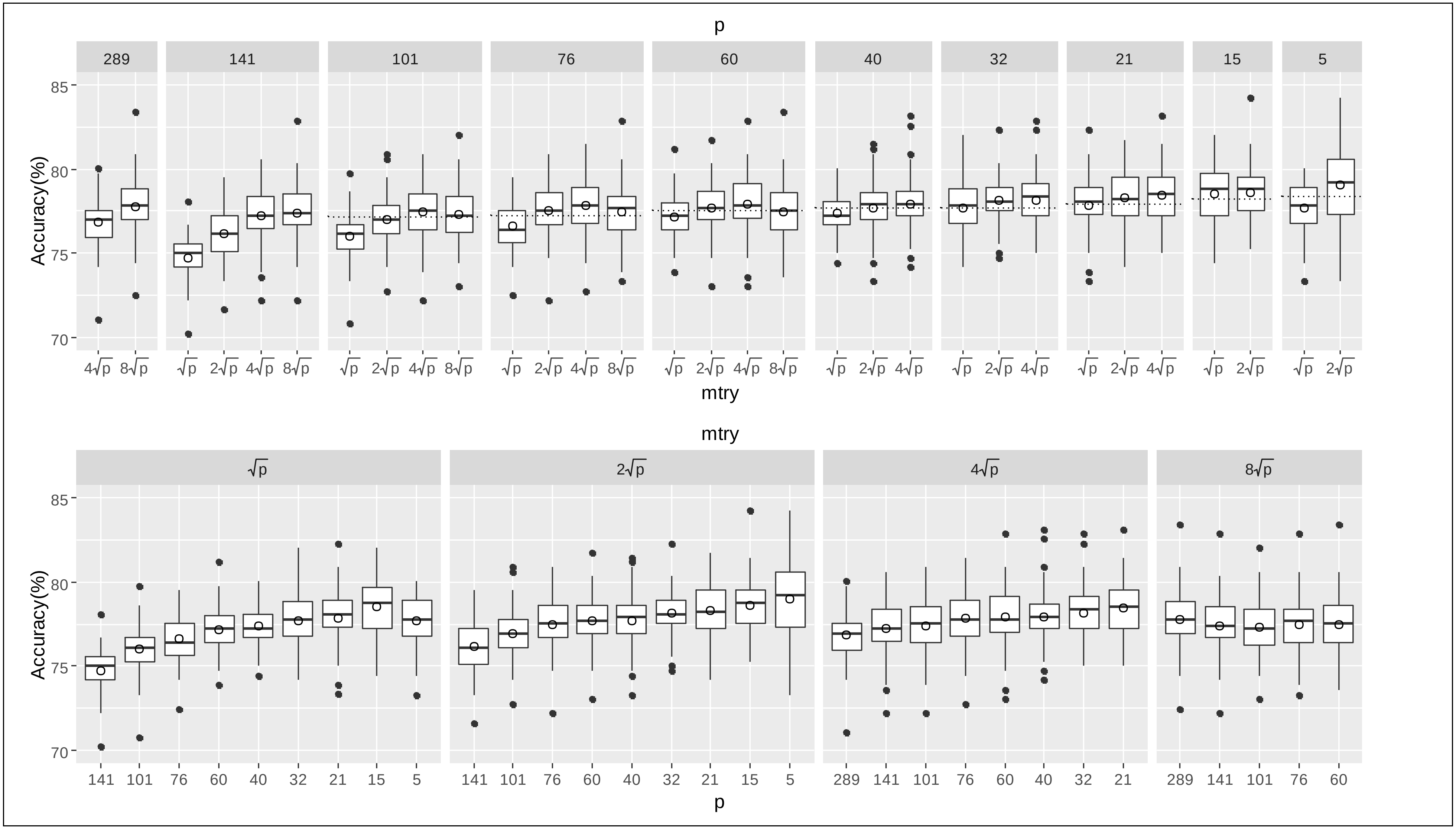
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model; Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model.

Figure 7.9. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicide attempts in Canadian general population.



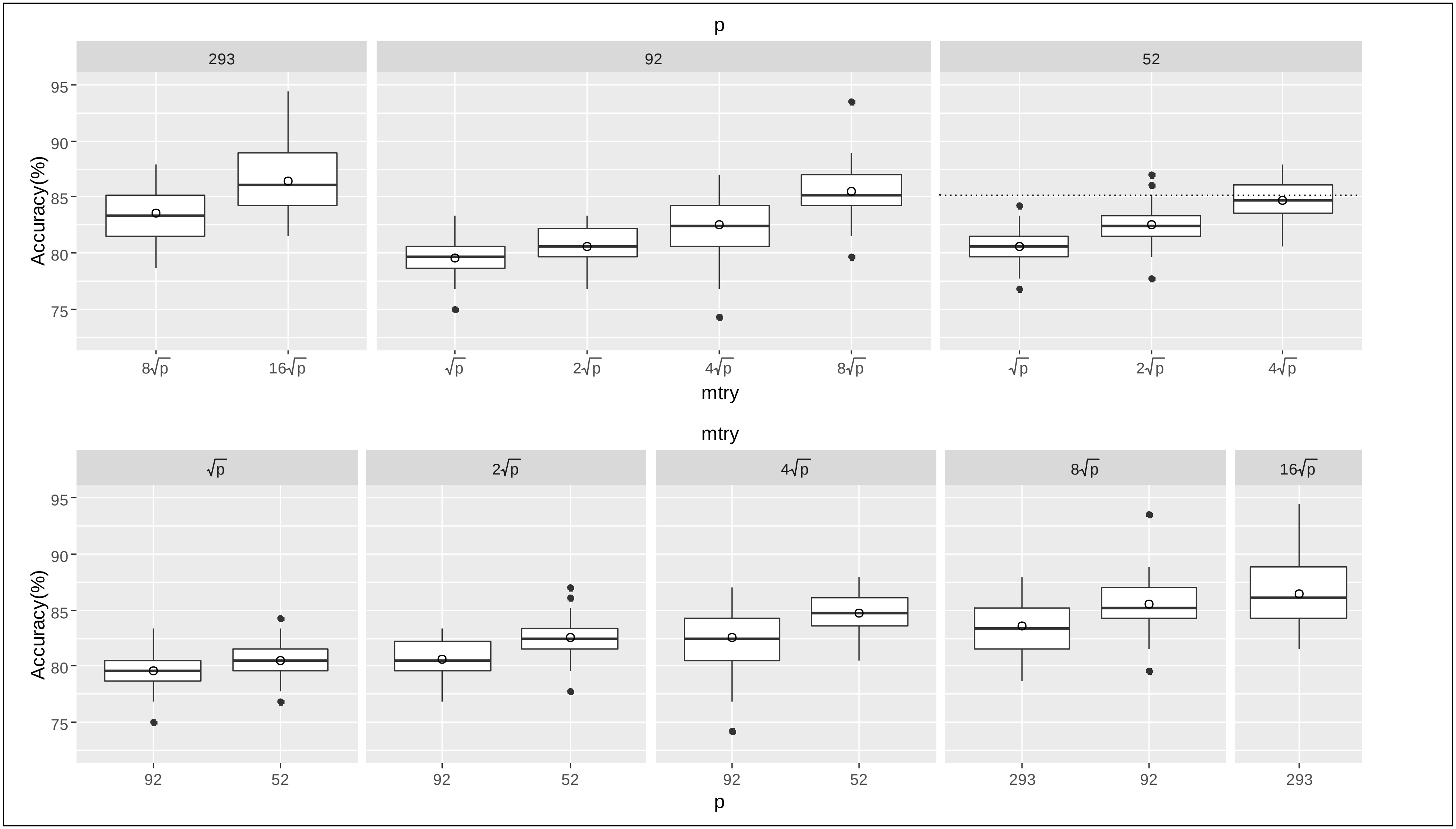
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model; Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model

Figure 7.10. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicidal ideation in Canadian Armed Forces.



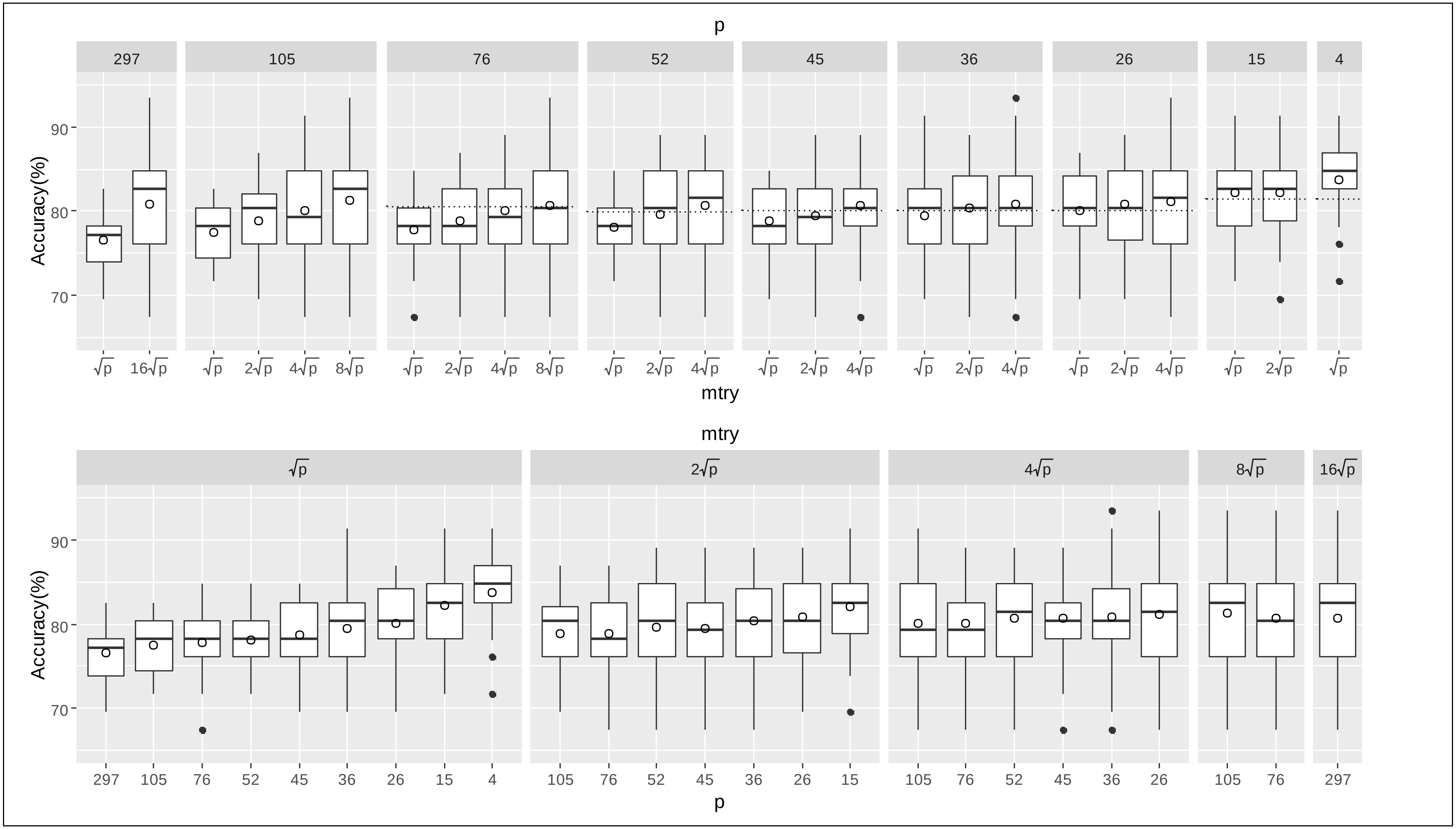
[1]Averaged over 50 iterations; [2]Including all p predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model.

Figure 7.11. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicide plans in Canadian Armed Forces.



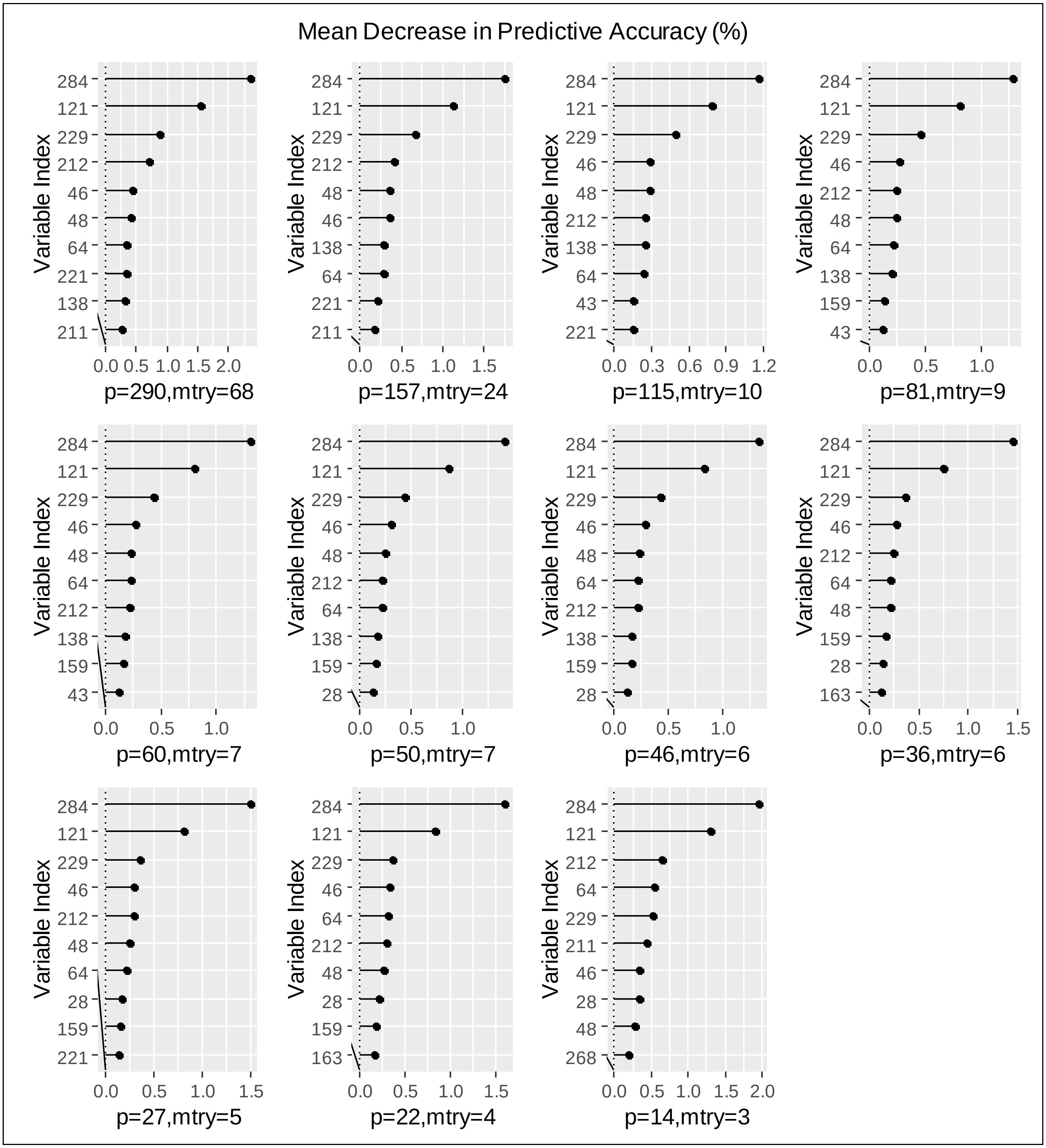
[1]Averaged over 50 iterations; [2]Including a subset of predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model; Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model.

Figure 7.12. Box plots of predictive accuracies[1] for reduced[2] conditional random forests models of suicide attempts in Canadian Armed Forces.



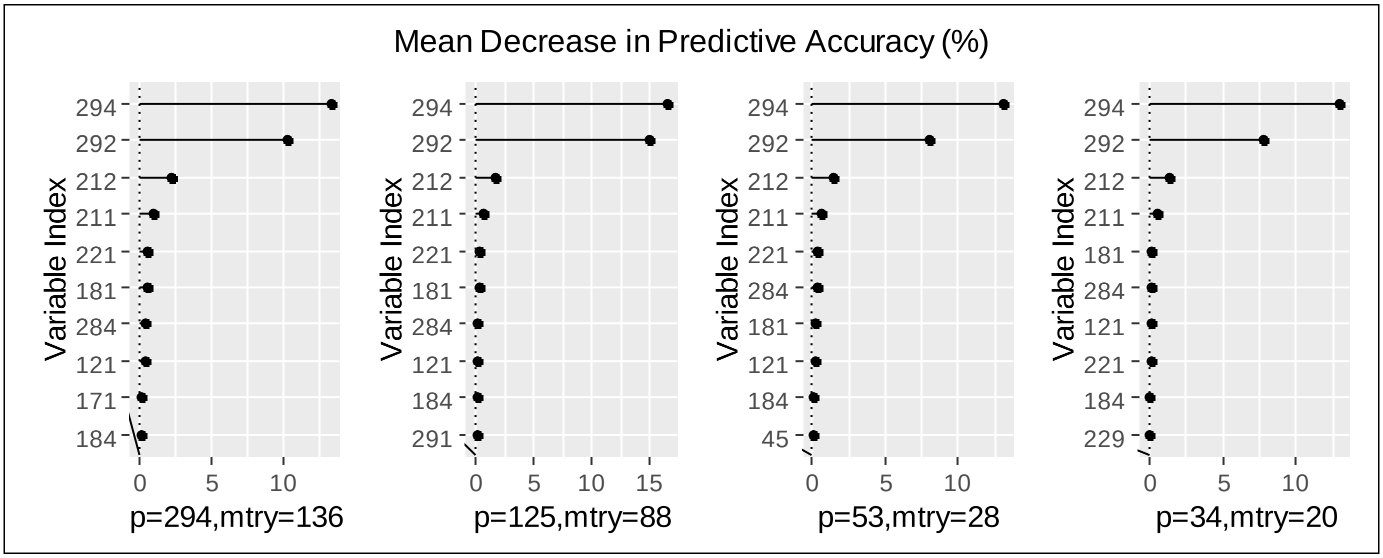
[1]Averaged over 50 iterations; [2]Including a subset of predictors; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; p= number of variables in the model; Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the previous best reduced model.

Figure 7.13. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicidal ideation in Canadian general population



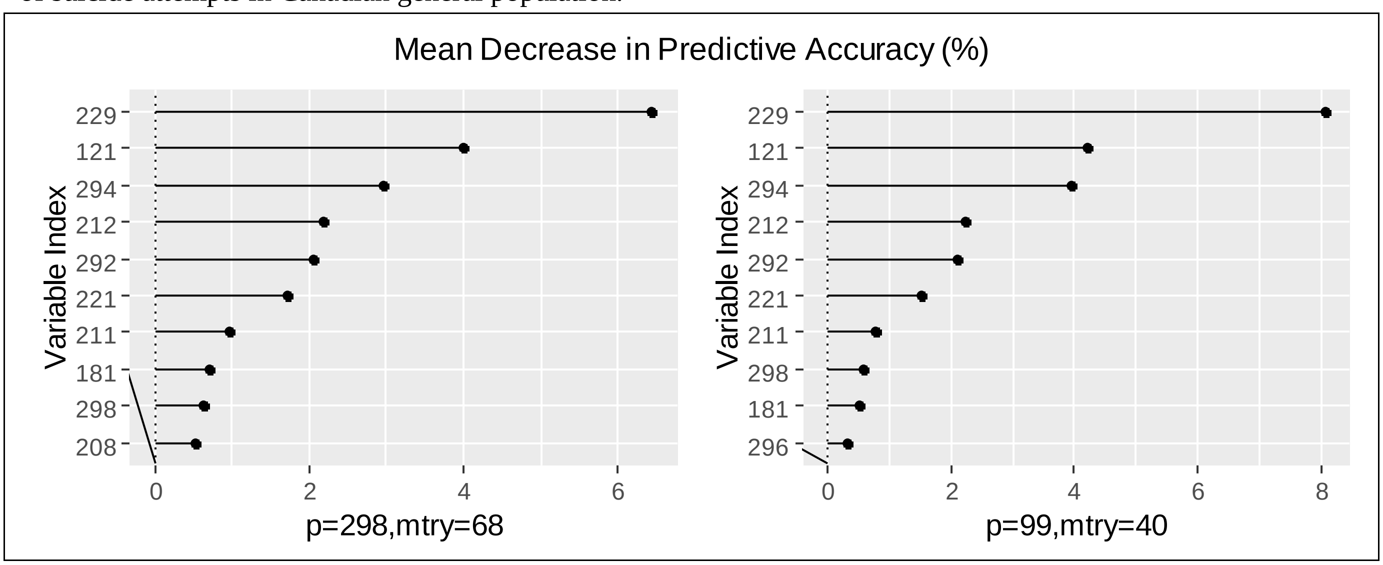
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=6260, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.14. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicide plans in Canadian general population.



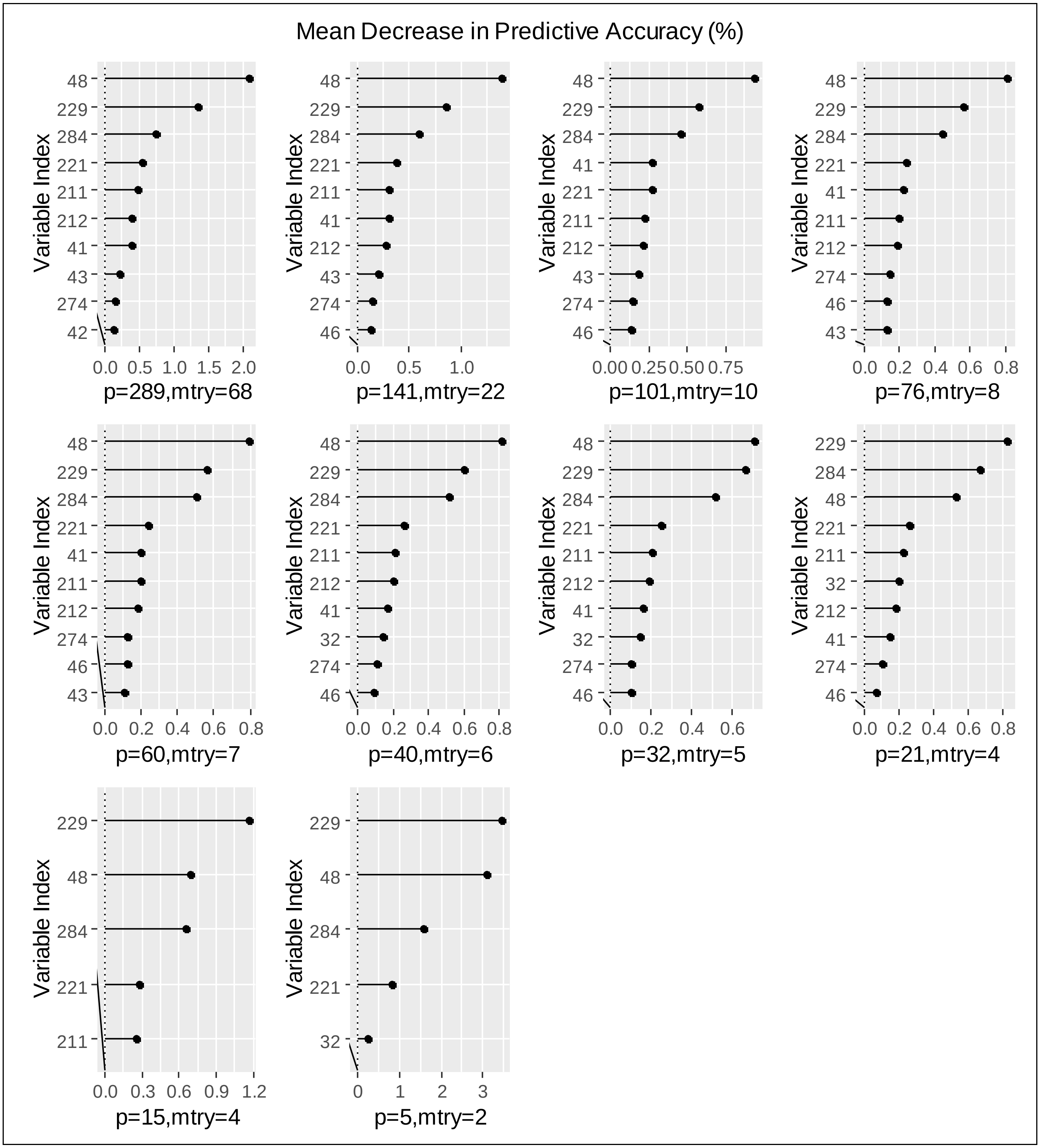
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=12518, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.15. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicide attempts in Canadian general population.



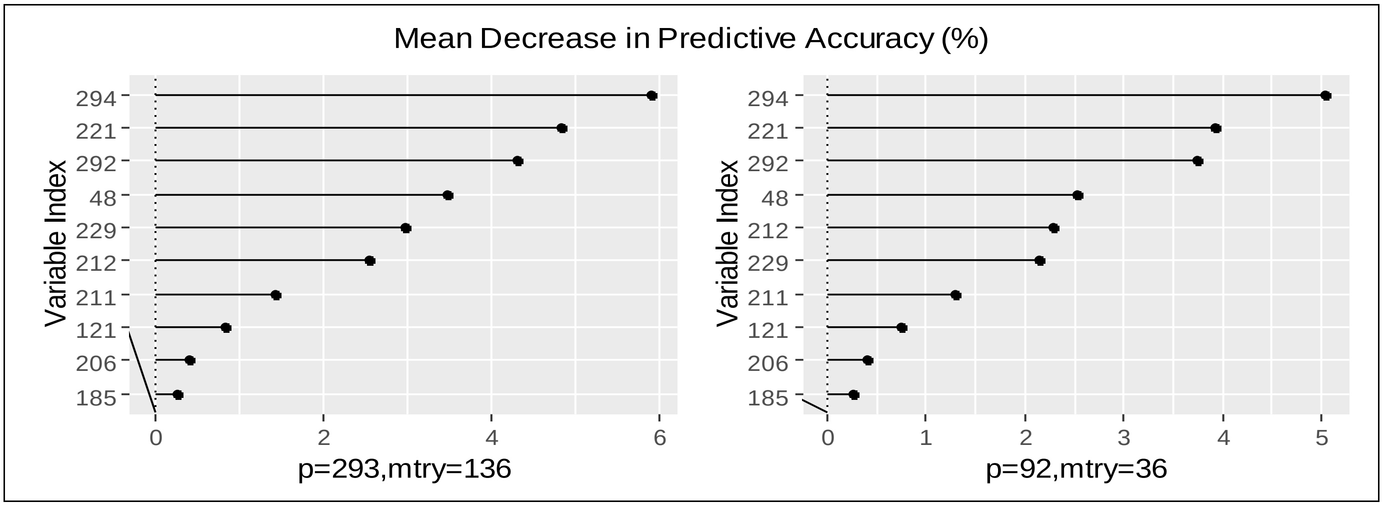
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=12518, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.16. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicidal ideation in Canadian Armed Forces.



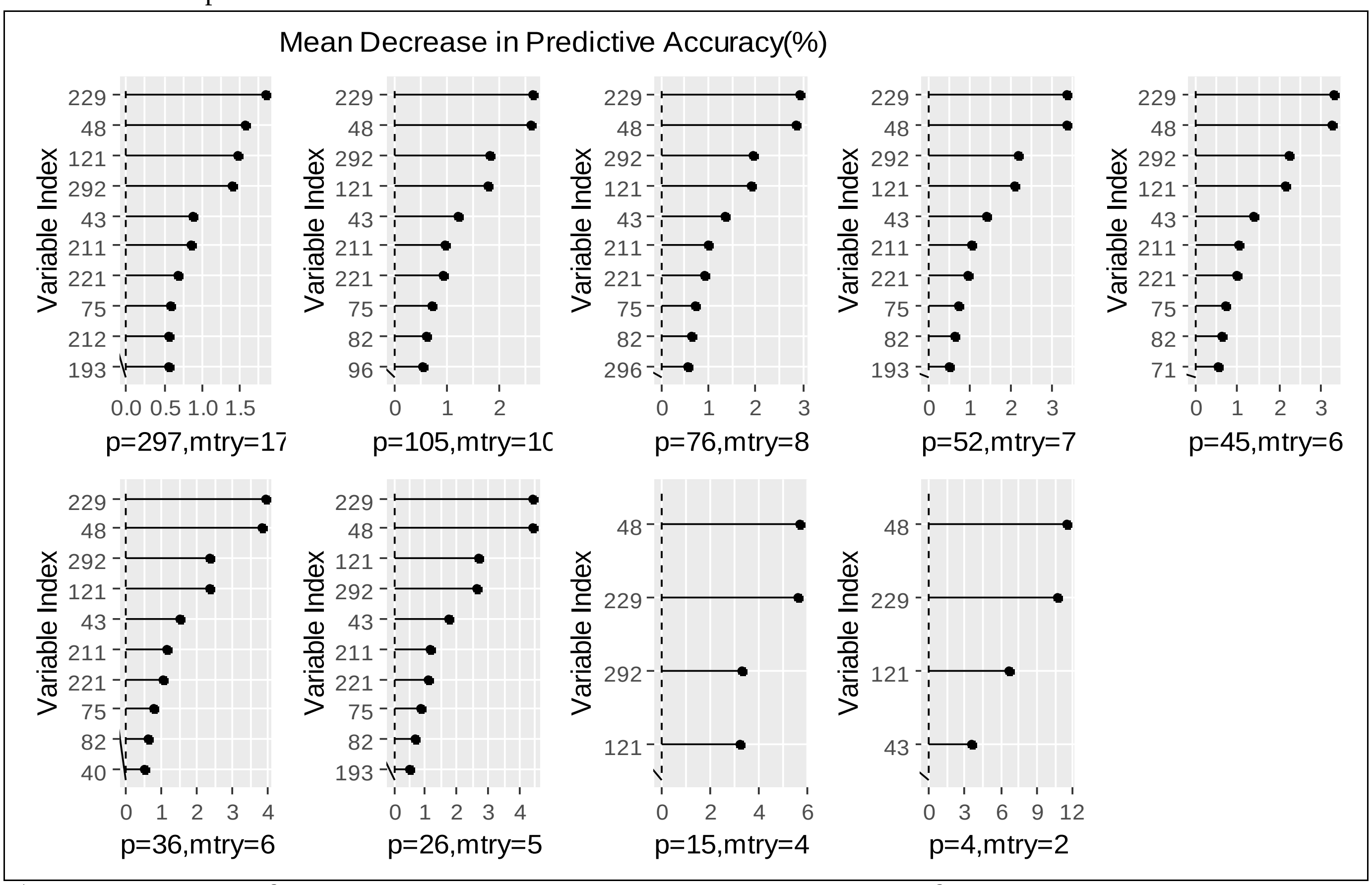
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=4077, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.17. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicide plans in Canadian Armed Forces.



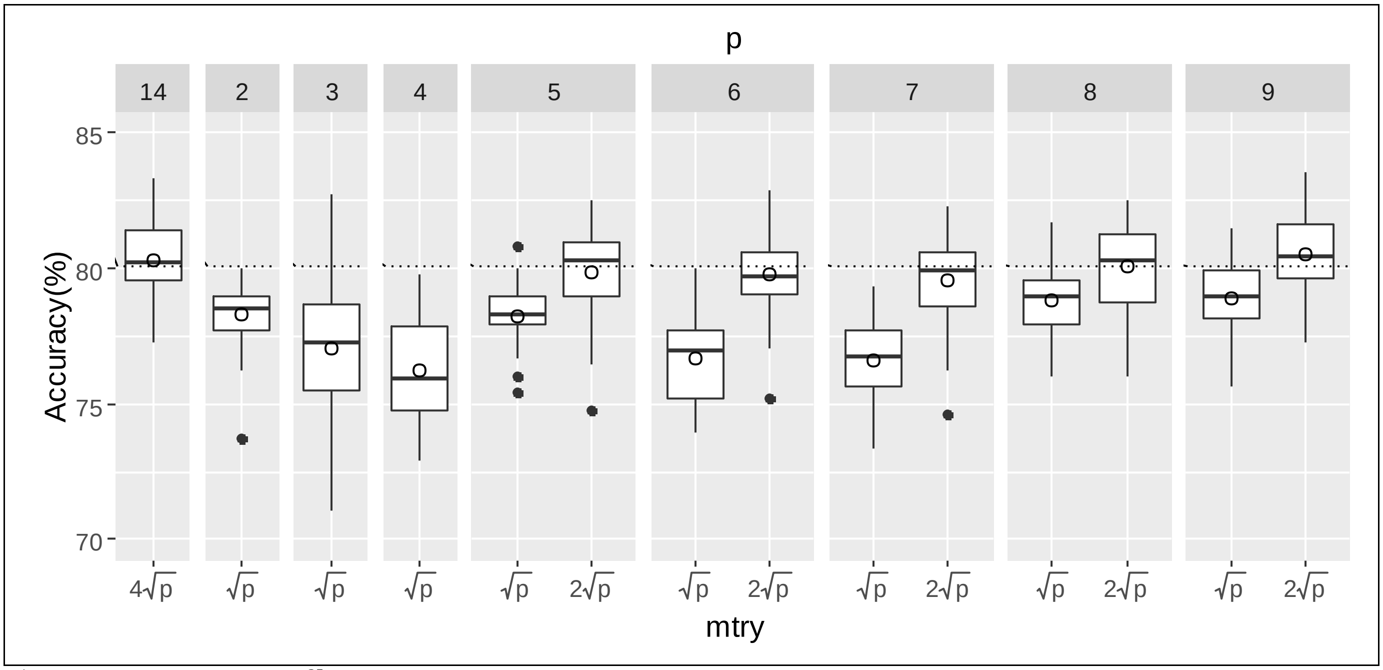
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=4076, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.18. Permutation accuracy importance scores[1,2] for predictors used in conditional random forests models[3] of suicide attempts in Canadian Armed Forces.



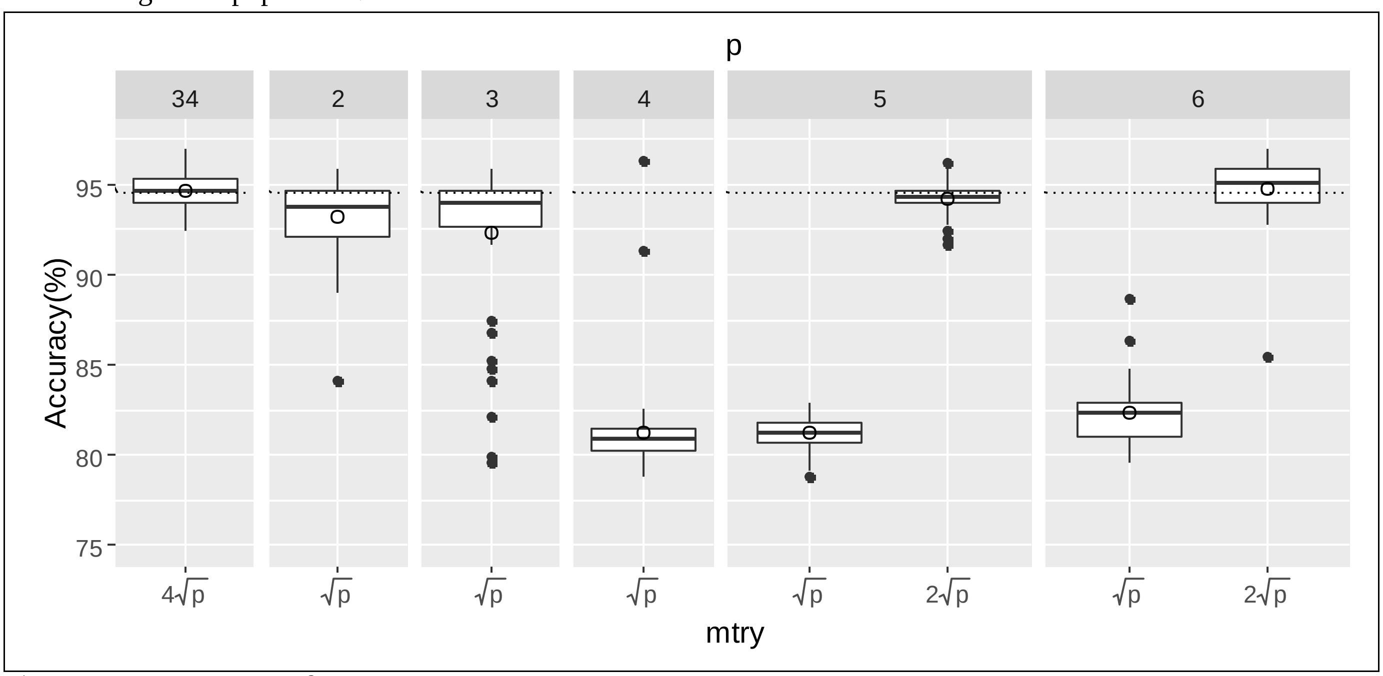
[1]Top 10 ranked variables; [2]Mean decrease in predictive accuracy is averaged over 50 iterations; [2]n_{train}=8153, *ntree*=500; *p*=number of variables in a classifier; *mtry*=number of randomly chosen variables at each split; *ntree*=number of trees grown; n_{train}=number of observations used for training the classifier.

Figure 7.19. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicidal ideation in Canadian general population.



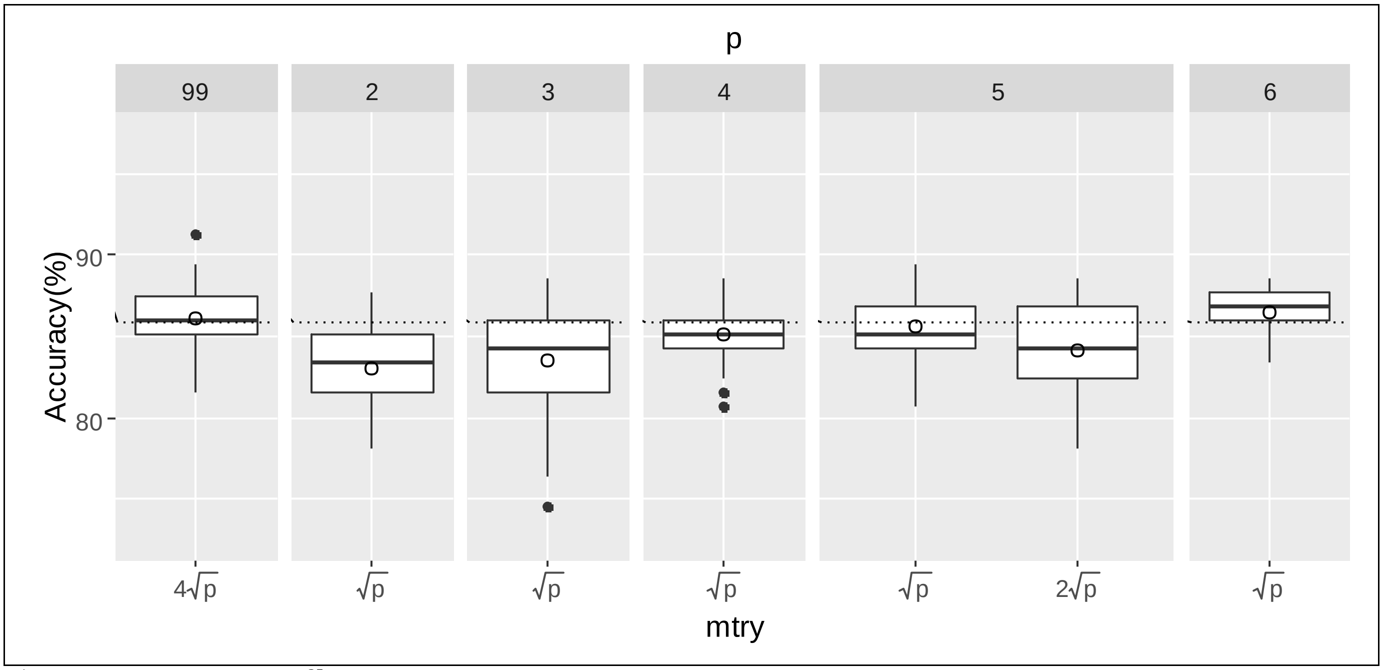
[1]Averaged over 50 iterations; [25]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (p=14); [3]n_{train}=6260, *ntree*=500; p= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.20. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicide plans in Canadian general population.



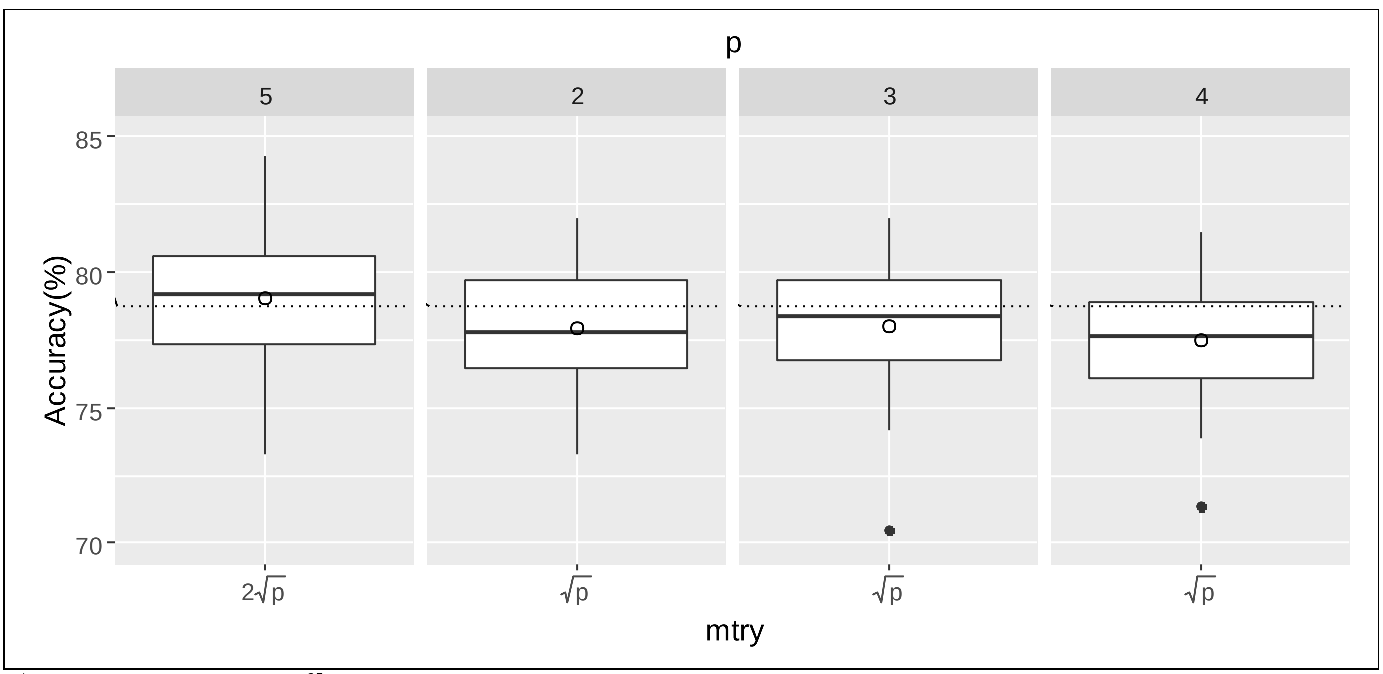
[1]Averaged over 50 iterations; [2]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (p=34); [3]n_{train}=12518, *ntree*=500; p= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.21. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicide attempts in Canadian general population.



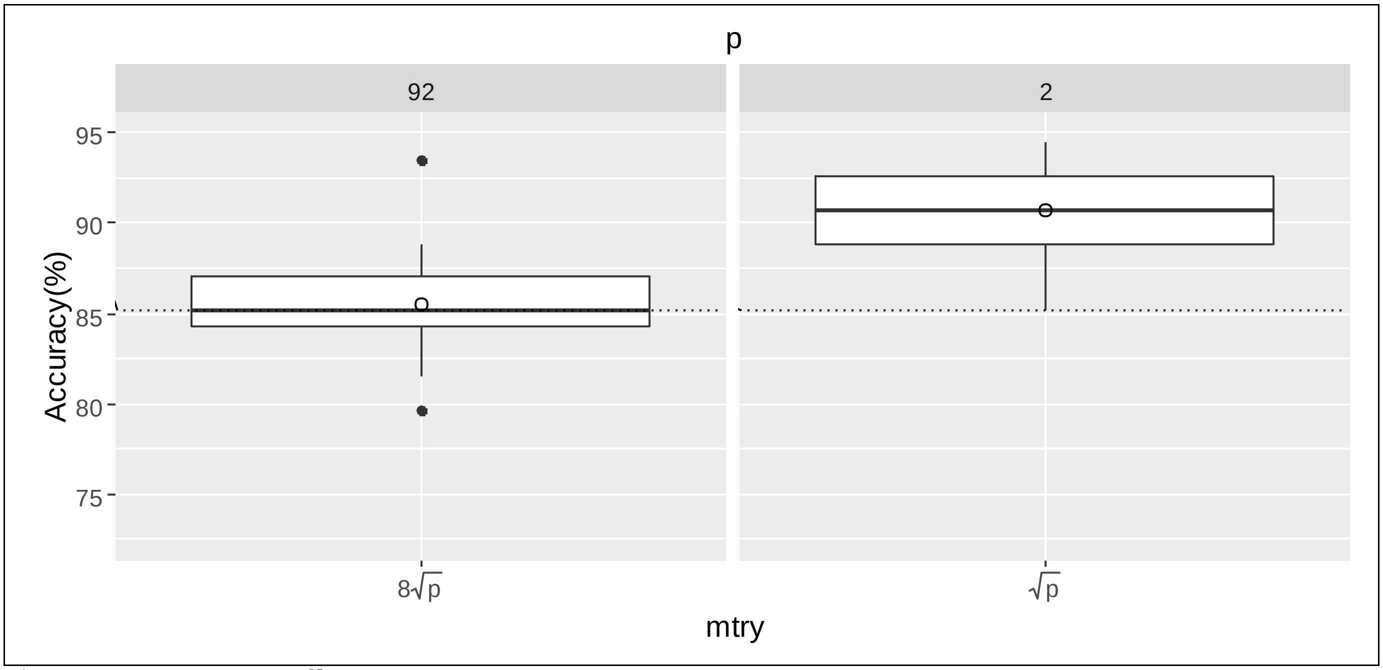
[1]Averaged over 50 iterations; [25]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (*p*=99); [3]n_{train}=12518, *ntree*=500; *p*= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.22. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicidal ideation in Canadian Armed Forces.



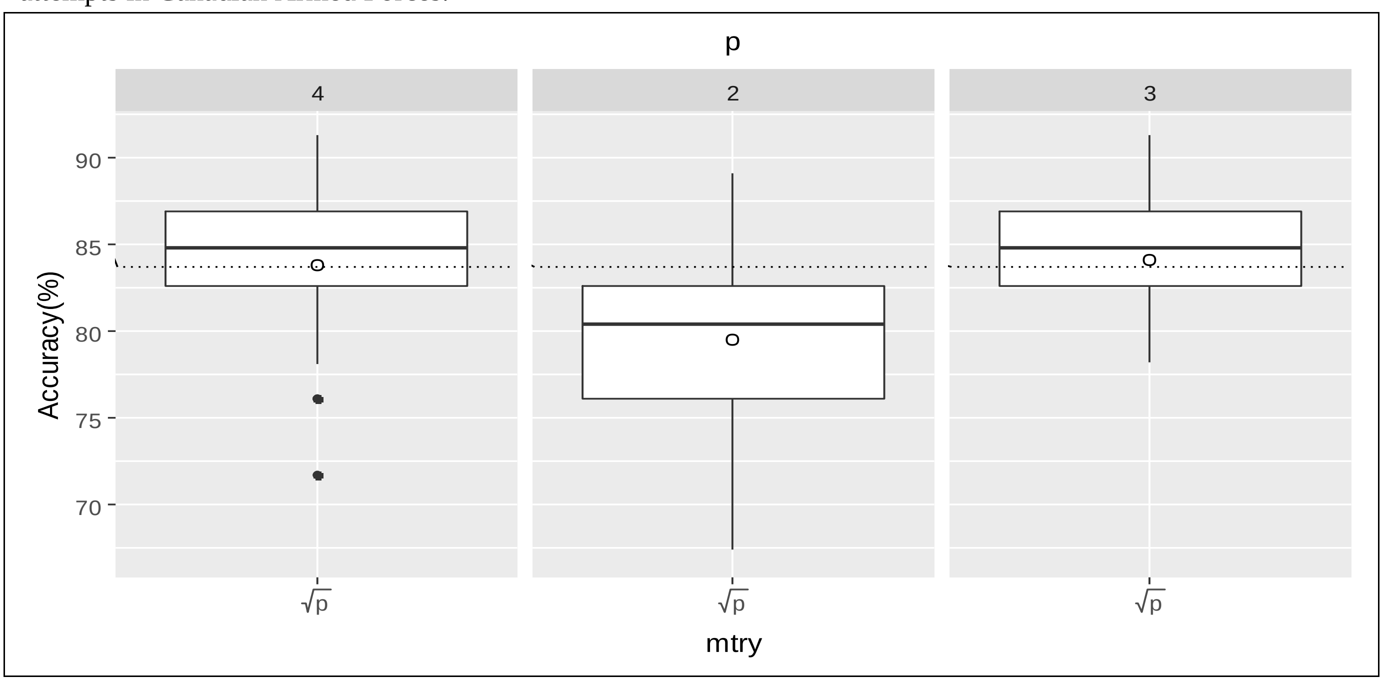
[1]Averaged over 50 iterations; [25]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (*p*=5); [3]n_{train}=4077, *ntree*=500; *p*= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.23. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicide plans in Canadian Armed Forces.



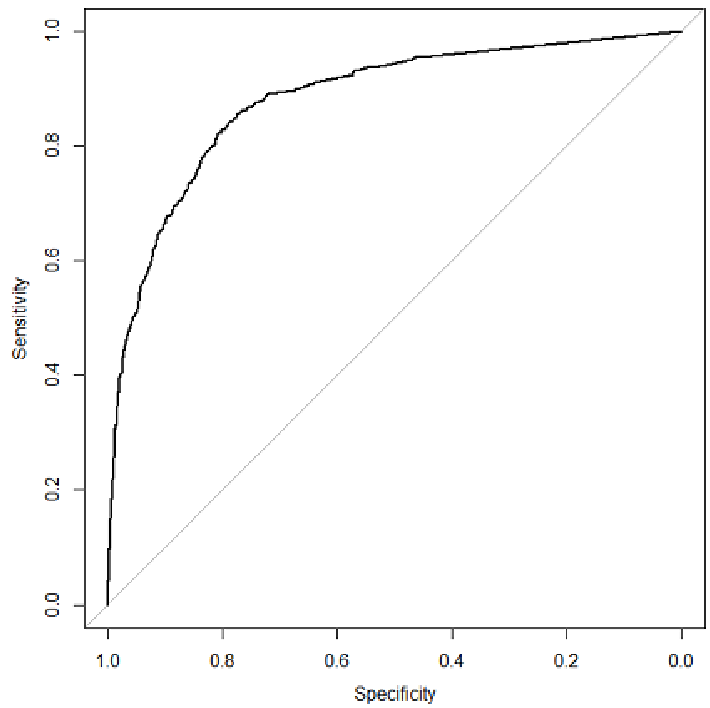
[1]Averaged over 50 iterations; [25]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (*p*=5); [3]n_{train}=4076, *ntree*=500; *p*= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.23. Box plots of predictive accuracies[1] for nested[2] conditional random forests models[3] of suicide attempts in Canadian Armed Forces.



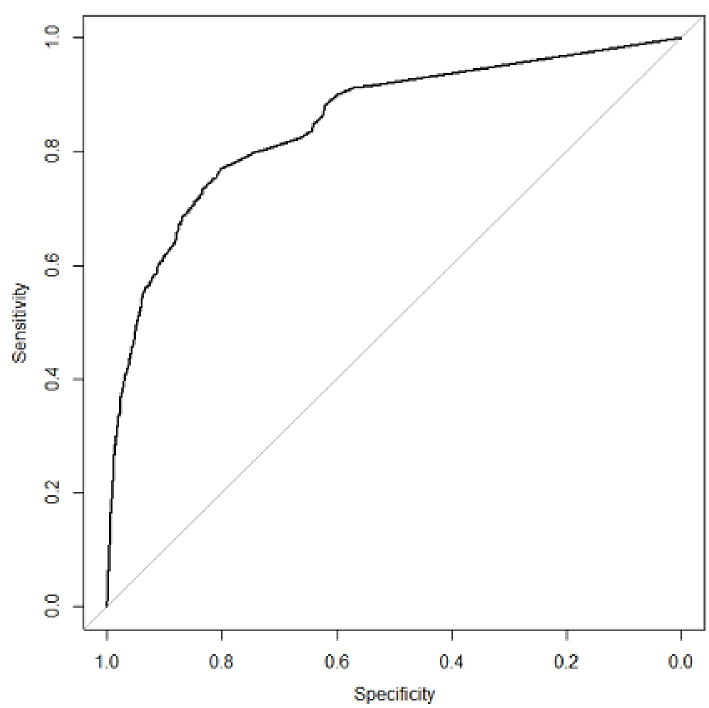
[1]Averaged over 50 iterations; [25]All nested models are models within the best reduced model, whose predictors are added, one at a time, according to their variable importance scores. Dotted line corresponds to a threshold value that is 1 standard error less than the predictive accuracy of the best reduced model (*p*=4); [3]n_{train}=8153, *ntree*=500; *p*= number of variables in the model; *mtry*=number of randomly chosen predictors at each split; *ntree*=number of trees grown; n_{train}=number of observations used to train the classifier.

Figure 7.25. Receiver operating characteristic curve for a conditional random forests model[1] of suicidal ideation trained on CCHS-MH sample and tested on CFMHS sample (AUC=0.8812)



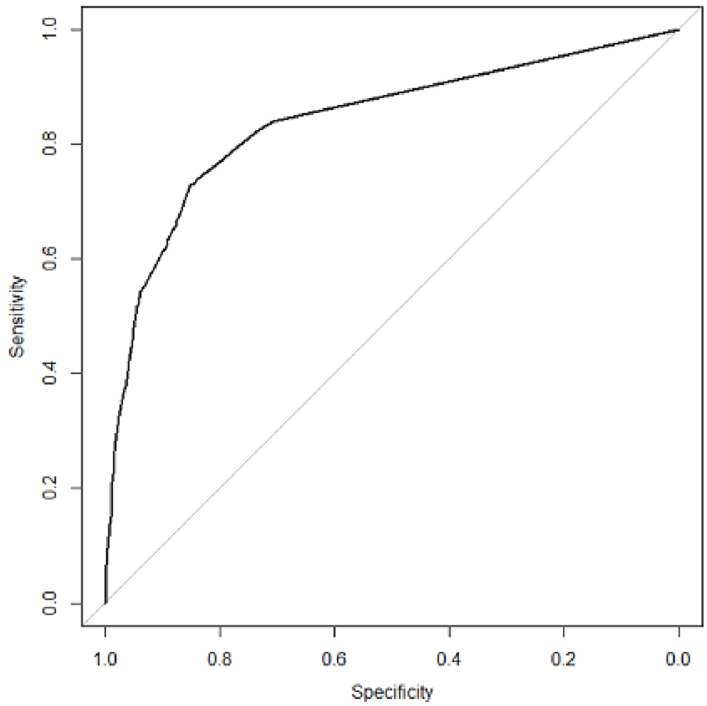
[1]n_{train}=25034; n_{test}=8152; p=9; mtry=6; ntree=500.

Figure 7.26. Receiver operating characteristic curve for a conditional random forests model[1] of suicidal ideation trained on half of CCHS-MH sample and tested on the other half of CCHS-MH sample (AUC=0.8505)



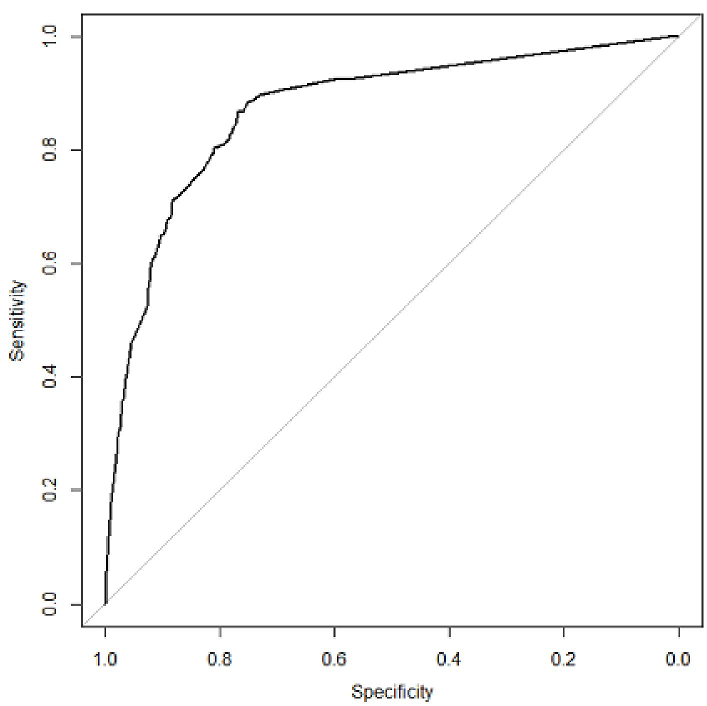
[1]n_{train}=12519; n_{test}=12591; p=9; mtry=6; ntree=500.

Figure 7.27. Receiver operating characteristic curve for a conditional random forests model[1] of suicidal ideation trained on CFMHS sample and tested on CCHS-MH sample (AUC=0.8391)



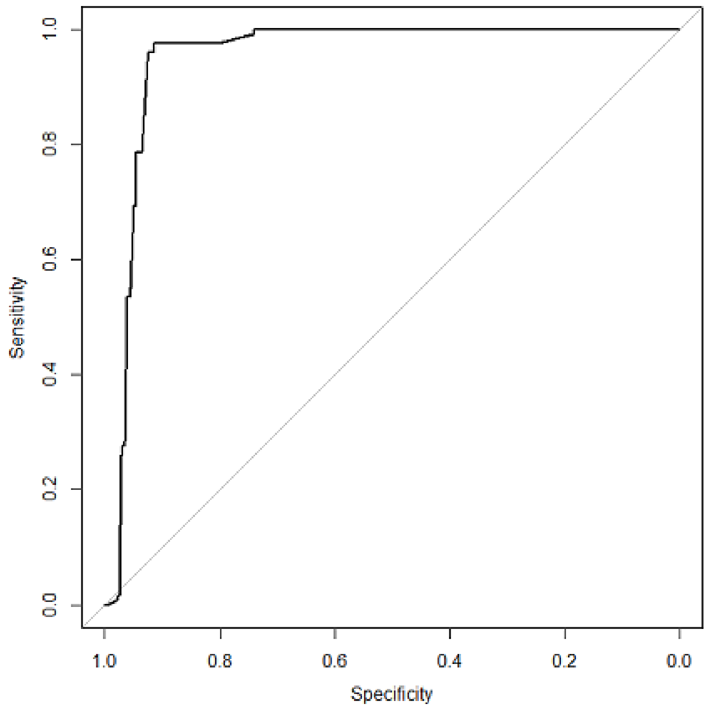
[1] n_{train}=8152; n_{test}=25034; p=5; mtry=4; ntree=500.

Figure 7.28. Receiver operating characteristic curve for a conditional random forests model[1] of suicidal ideation trained on half of CFMHS sample and tested on the other half of CFMHS sample (AUC=0.8728)



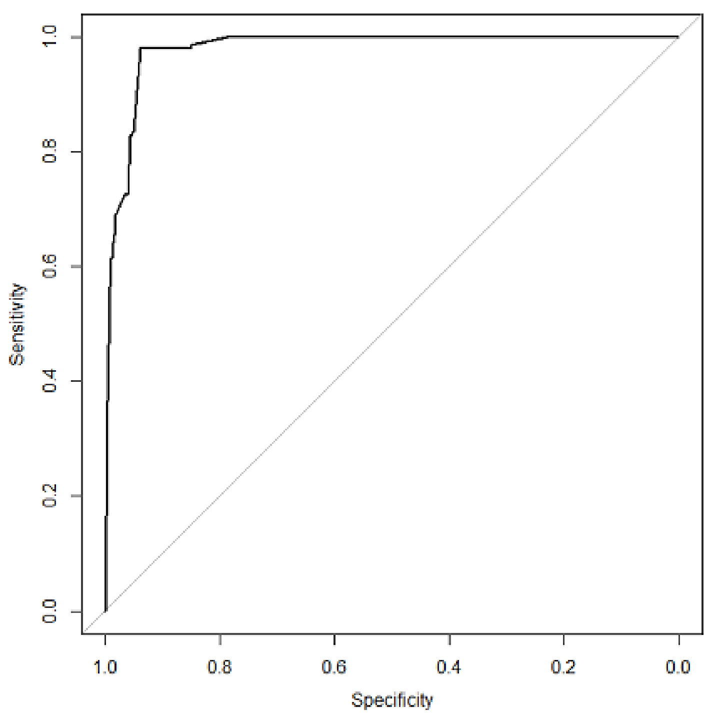
[1] n_{train}=8152; n_{test}=25034; p=5; mtry=4; ntree=500.

Figure 7.29. Receiver operating characteristic curve for a conditional random forests model[1] of suicide plans trained on CCHS-MH sample and tested on CFMHS sample (AUC=0.9520)



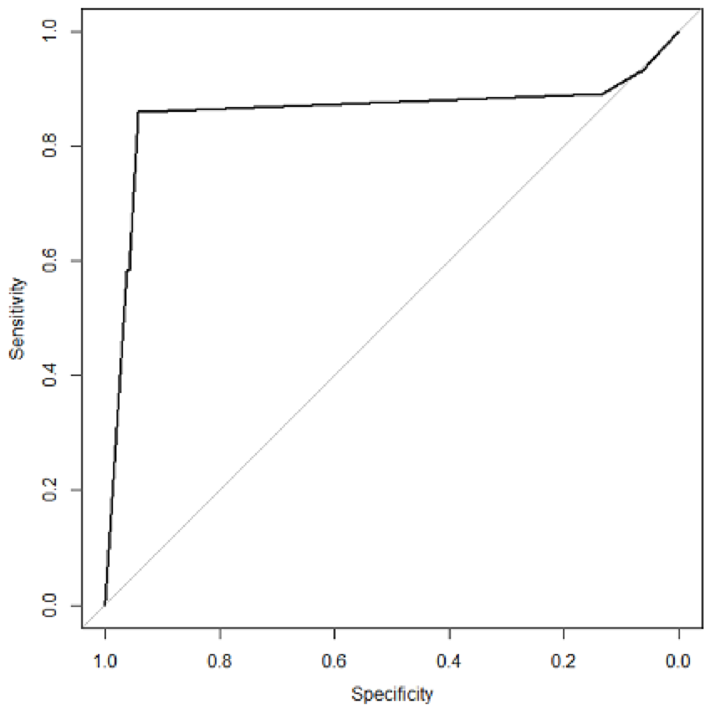
[1]n_{train}=25034; n_{test}=8152; p=6; mtry=4; ntree=500.

Figure 7.30. Receiver operating characteristic curve for a conditional random forests model[1] of suicide plans trained on half of CCHS-MH sample and tested on the other half of CCHS-MH sample (AUC=0.9793)



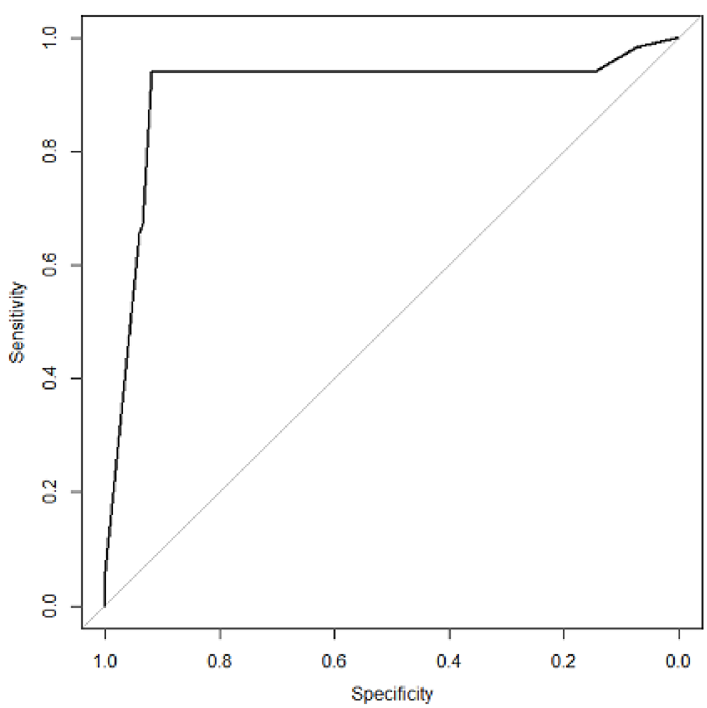
[1]n_{train}=12517; n_{test}=12517; p=6; mtry=4; ntree=500.

Figure 7.31. Receiver operating characteristic curve for a conditional random forests model[1] of suicide plans trained on CFMHS sample and tested on CCHS-MH sample (AUC=0.8580)



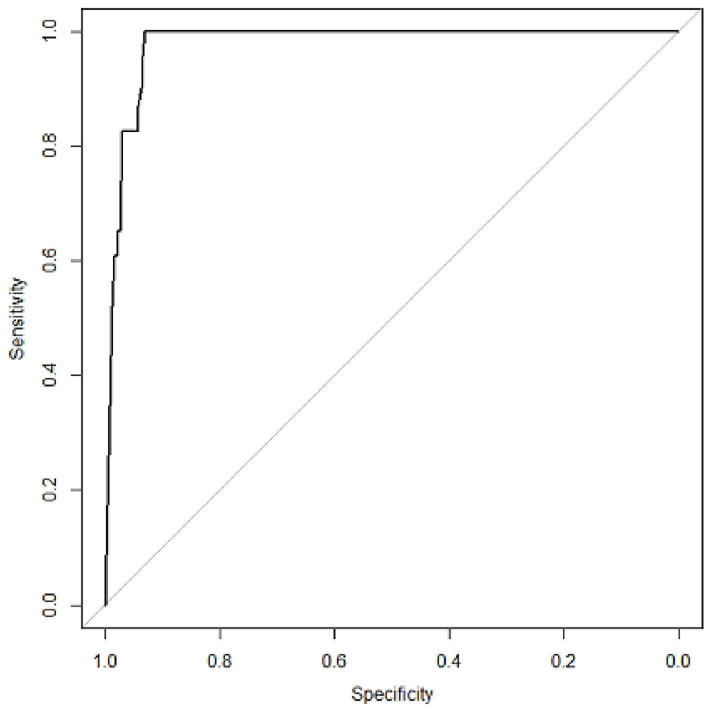
[1] n_{train}=8152; n_{test}=25034; p=2; mtry=1; ntree=500.

Figure 7.32. Receiver operating characteristic curve for a conditional random forests model[1] of suicide plans trained on half of CFMHS sample and tested on the other half of CFMHS sample (AUC=0.9072)



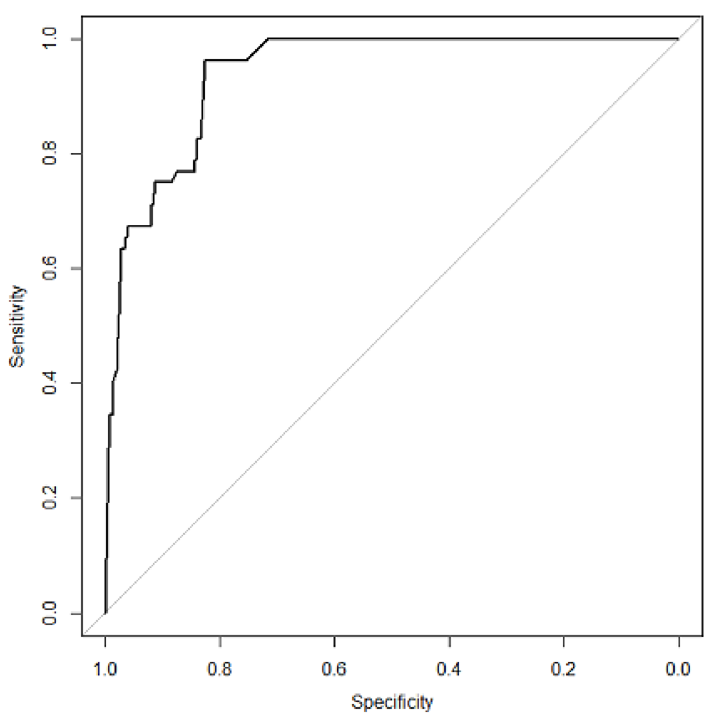
[1] n_{train}=4076; n_{test}=4076; p=2; mtry=1; ntree=500.

Figure 7.33. Receiver operating characteristic curve for a conditional random forests model[1] of suicide attempts trained on CCHS-MH sample and tested on CFMHS sample (AUC=0.9786)



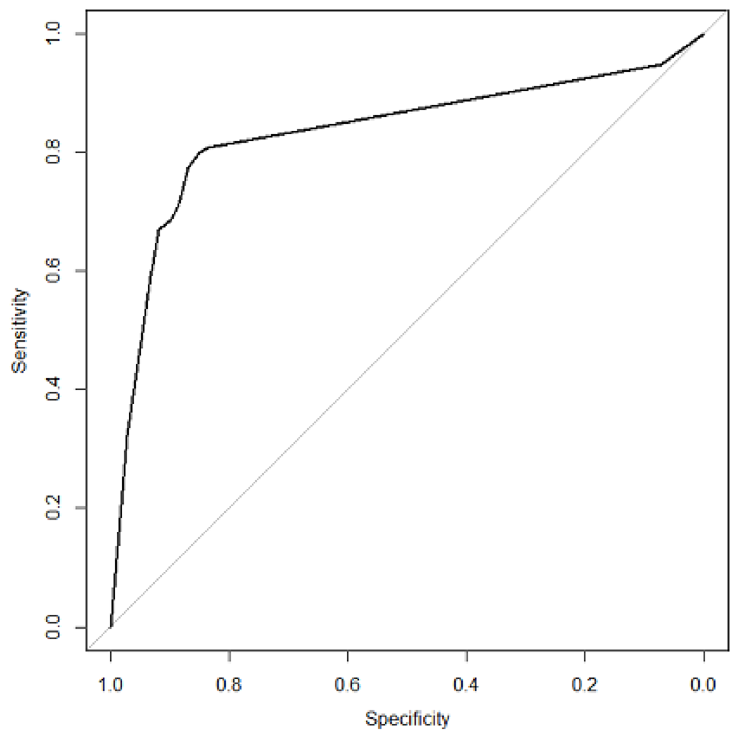
[1]n_{train}=25036; n_{test}=8153; p=6; mtry=4; ntree=500.

Figure 7.34. Receiver operating characteristic curve for a conditional random forests model[1] of suicide attempts trained on half of CCHS-MH sample and tested on the other half of CCHS-MH sample (AUC=0.9404)



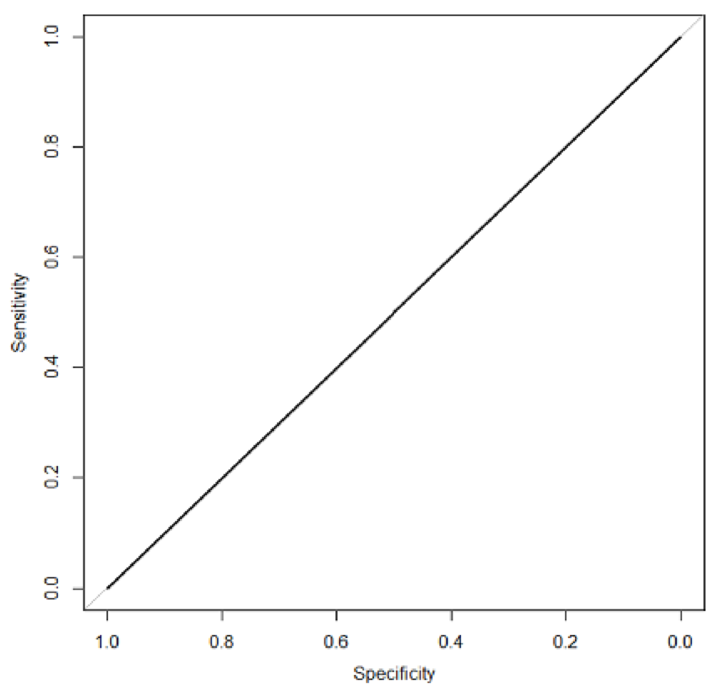
[1]n_{train}=12518; n_{test}=12518; p=6; mtry=4; ntree=500.

Figure 7.35. Receiver operating characteristic curve for a conditional random forests model[1] of suicide attempts trained on CFMHS sample and tested on CCHS-MH sample (AUC=0.8334)



[1]n_{train}=8153; n_{test}=25036; p=3; mtry=2; ntree=500.

Figure 7.36. Receiver operating characteristic curve for a conditional random forests model[1] of suicide attempts trained on half of CFMHS sample and tested on the other half of CFMHS sample (AUC=0.5)



[1]n_{train}=4077; n_{test}=4076; p=3; mtry=2; ntree=500.

8. References

1. Mann JJ. Neurobiology of suicidal behaviour. Nature Reviews: Neuroscience. 2003; 4(10): 819-826.

2. World Health Organization. Preventing suicide: a global imperative. Geneva, Switzerland; 2014.

3. Navaneelan T. Suicide rates: An overview. Health at a Glance. 2012; 82: 3-11.

4. Public Health Agency of Canada. Working together to prevent suicide in Canada: the federal framework for suicide prevention. 2016. Available at: https://www.canada.ca/en/public-health/services/publications/healthy-living/suicide-prevention-framework.html.

5. Statistics Canada. Table 102-0551 - Deaths and mortality rate, by selected grouped causes, age group and sex, Canada, annual, CANSIM database. 2017; Accessed: July 27, 2017.

6. Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. BMC Psychiatry. 2004;4(37):1-11.

7. Joiner TE, Buchman-Schmitt JM, Chu C. Do undiagnosed suicide decedents have symptoms of a mental disorder? Journal of Clinical Psychology. 2017; doi:10.1002/jclp.22498.

8. Rusu C, Zamorski MA, Boulos D, Garber BG. Prevalence comparison of past-year mental disorders and suicidal behaviours in the Canadian Armed Forces and the Canadian general population. The Canadian Journal of Psychiatry. 2016;61(1_suppl):46S-55S.

9. Borges G, Angst J, Nock MK, Ruscio AM, Kessler RC. Risk factors for the incidence and persistence of suicide-related outcomes: a 10-year follow-up study using the National Comorbidity Surveys. Journal of Affective Disorders. 2008;105(1-3):25-33.

10. Statistics Canada. Canadian Forces Cancer and Mortality Survey: Causes of Death. 2011. Available at: http://www.statcan.gc.ca/pub/82-584-x/2011001/tbl/82-584-02-tbl-eng.htm.

11. Mackenzie L. Canadian Forces: holding the line on mental health. The Globe and Mail. February 14, 2014. Available at: https://www.theglobeandmail.com/opinion/canadian-forces-holding-the-line-on-mental-health/article16892831.

12. Zamorski MA. Report of the Canadian Forces Expert Panel on Suicide Prevention. Department of National Defence. Ottawa (ON); 2010.

13. Bogaert L, Whitehead J, Wiens M, Rolland E. Suicide in the Canadian Forces 1995 to 2012. Directorate of Force Health Protection. Surgeon General Report. Ottawa (ON); 2013.

14. Ramchand R, Acosta J, Burns RM, Jaycox LH, Pernin CG. The War Within: Preventing Suicide in the U.S. Military. Santa Monica, CA: The RAND Corporation, 2011.

15. Hyman J, Ireland R, Frost L, Cottrell L. Suicide incidence and risk factors in an active duty U.S. military population. American Journal of Public Health. 2012; 102(Suppl 1): S138-146.

16. Kuehn BM. Soldier suicide rates continue to rise: Military, scientist work to stem the tide. JAMA 2009; 301, 1112-1113.

17. Rolland-Harris E, Whitehead J, Matheson H, Zamorski MA. 2015 report on suicide mortality in the Canadian Armed Forces (1995-2014). National Defence, Directorate of Force Health Protection, Directorate of Mental Health. Surgeon General Report. Ottawa (ON); 2015.

18. Rolland-Harris E, Cyr E, Zamorski MA. 2016 report on suicide mortality in the Canadian Armed Forces (1995-2015). National Defence, Directorate of Force Health Protection, Directorate of Mental Health. Surgeon General Report. Ottawa (ON); 2016.

19. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiologic Reviews. 2008. 30(1): 133-154.

20. Nock MK, Deming CA, Fullerton CS, Gilman SE, Goldenberg M, Kessler RC, McCarroll JE, McLaughlin KA, Peterson C, Schoenbaum M, Stanley B, Ursano RJ. Suicide among soldiers: a review of psychosocial risk and protective factors. Psychiatry. 2013. 76(2): 97-125.

21. Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite S, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev. 2010; 117(2): 575-600.

22. Mościcki EK. Suicidal behaviors among adults. The Oxford Handbook of Suicide and Self-Injury. 2014;82.

23. Cavanagh JT, Carson AJ, Sharpe M, Lawrie SM. Psychological autopsy studies of suicide: a systematic review. Psychological Medicine. 2003; 33:395-405.

24. Nock MK, Dempsey CL, Aliaga PA, Heeringa SG, Kessler RC, Stein MB, Ursano RJ, Benedek D. Psychological autopsy study comparing suicide decedents, suicide ideators, and propensity score matched controls: results from the study to assess risk and resilience in service members (Army STARRS). Psychological Medicine. 2017;1-12.

25. Beautrais AL, Joyce PR, Mulder RT, Fergusson DM, Deavoll BJ, Nightingale SK. Prevalence and comorbidity of mental disorders in persons making serious suicide attempts: a case-control study. American Journal of Psychiatry. 1996;153:1009-1014.

26. Shaffer D, Gould MS, Fisher P, Trautman P, Moreau D, Kleinmen M, Flory M. Psychiatric diagnosis in child and adolescent suicide. Archives of General Psychiatry. 1996; 53(4): 339-348.

27. Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. British Journal of Psychiatry. 1997: 170: 205-228.

28. Kessler RC, Borges G, Walters EE. Prevalence of and risk factors for lifetime suicide attempts in the National Comorbidity Survey. Archives of General Psychiatry. 1999; 56(7): 617-626.

29. Mann JJ, Waternaux C, Haas GL, Malone KM. Toward a clinical model of suicidal behavior in psychiatric patients. American Journal of Psychiatry. 1999; 156: 181-189.

30. Wilcox HC, Conner KR, Caine ED. Association of alcohol and drug use disorders and completed suicide: an empirical review of cohort studies. Drug and Alcohol Dependence. 2004; 76 Suppl: S11-19.

31. Hawton K, Sutton L, Haw C, Sinclair J, Harris L. Suicide and attempted suicide in bipolar disorder: a systematic review of risk factors. Journal of Clinical Psychiatry. 2005; 66(6): 693-704.

32. Sareen J, Houlahan T, Cox BJ, Asmundson GJ. Anxiety disorders associated with suicidal ideation and suicide attempts in the National Comorbidity Survey. Journal of Nervous and Mental Disease. 2005; 193: 450-454.

33. Sareen J, Cox BJ, Afifi TO, de Graaf R, Asmundson GJ, ten Have M, Stein MB. Anxiety disorders and risk for suicidal ideation and suicide attempts: a population-based longitudinal study of adults. Archives of General Psychiatry. 2005; 62(11): 1249-1257.

34. Sareen J, Cox BJ, Stein MB, Afifi TO, Stein MB, Belik SL, Meadows G, Asmundson GJ. Combat and peacekeeping operations in relation to prevalence of mental disorders and perceived need for mental health care. Archives of General Psychiatry. 2007; 64(7): 843-852.

35. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, Bruffaerts R, Chiu WT, de Girolamo G, Gluzman S, de Graaf R, Gureje O, Haro JM, Huang Y, Karam E, Kessler RC, Lepine JP, Levinson D, Medina-More ME, Ono Y, Posada-Villa J, Williams D. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. British Journal of Psychiatry 2008;192:98-105.

36. Nock MK, Hwang I, Sampson N, Kessler RC, Angermeyer M, Beautrais A, Borges G, Bromet E, Bruffaerts R, de Girolamo G, de Graaf R, Florescu S, Gureje O, Haro JM, Hu C, Huang Y, Karam EG, Kawakami N, Kovess V, Levinson D, Posada-Villa J, Sagar R, Tomov T, Viana MC, Williams DR. Cross-national analysis of the association between traumatic events and suicidal behavior: findings from the WHO World Mental Health Surveys. PLoS Med. 2009; 6(8):e1000123.

37. Nock MK, Hwang I, Sampson NA, Kessler RC. Mental disorders, comorbidity and suicidal behavior: results from the National Comorbidity Survey Replication. Molecular Psychiatry. 2010; 15(8): 868-876.

38. Belik SL, Stein MB, Asmundson GJ, Sareen J. Are Canadian soldiers more likely to have suicidal ideation and suicide attempts than Canadian civilians? American Journal of Epidemiology. 2010;172(11):1250-1258.

39. Borges G, Nock MK, Abad JNM, Hwang I, Sampson NA et al. Twelve month prevalence of and risk factors for suicide attempts in the WHO World Mental Health Survey. Journal of Clinical Psychiatry. 2010;71(12):1617-1628.

40. Nock MK, Stein MB, Heeringa SG, Ursano RJ, Colpe LJ, Fullerton CS, Hwang I, Naifeh JA, Sampson NA, Schoenbaum M, Zaslavsky AM, Kessler RC. Prevalence and correlates of suicidal behavior among soldiers. Results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). JAMA Psychiatry. 2014;71(5):514-522.

41. Hawton K, Houston K, Haw C, Townsend E, Harris L. Comorbidity of axis I and axis II disorders in patients who attempted suicide. American Journal of Psychiatry. 2003; 160(8): 1494-1500.

42. Blasco-Fontecilla H, Rodrigo-Yanguas M, Giner L, Lobato-Rodriguez, MJ, de Leon J. Patterns of comorbidity of suicide attempters: an update. Current Psychiatry Reports. 2016;18(1):93.

43. Fawcett J, Scheftner WA, Fogg L, Clark DC, Young MA, Hedeker D, Gibbons R. Time-related predictors of suicide in major affective disorder. American Journal of Psychiatry. 1990; 147(9); 1189-1194.

44. Brent Da, Baugher M, Bridge J, Chen T, Chiapetta L. Age- and sex-related risk factors for adolescent suicide. Journal of the American Academy of Child and Adolescent Psychiatry. 1999; 38(12): 1497-1505.

45. Colman I, Newman SC, Schopflocher D, Bland RC, Dyck RJ. A multivariate study of predictors of repeat parasuicide. Acta Psychiatrica Scandinavica. 2004;109(4):306-12.

46. Coryell W, Young EA. Clinical predictors of suicide in primary major depressive disorder. Journal of Clinical Psychiatry. 2005; 66(4): 412-417.

47. Borges G, Angst J, Nock MK, Ruscio AM, Walters EE, Kessler RC. Risk factors for twelve-month suicide attempts in the National Comorbidity Survey Replication (NCS-R). Psychological Medicine. 2006;36(12):1747-1757.

48. Haw C, Bergen H, Casey D, Hawton K. Repetition of deliberate self-harm: a study of the characteristics and subsequent deaths in patients presenting to a general hospital according to extent of repetition. Suicide and Life-Threatening Behavior. 2007; 37(4): 379-396.

49. Suominen K, Isometsa E, Suokas J, Haukka J, Achte K, Lonnqvist J. Completed suicide after a suicide attempt: a 37-year follow-up study. American Journal of Psychiatry. 2004; 161: 563-564.

50. Bryan CJ, Bryan AO, Ray-Sannerud BN, Etienne N, Morrow CE. Suicide attempts before joining the military increase risk for suicide attempts and severity of suicidal ideation among military personnel and veterans. Comprehensive Psychiatry. 2013;55(3):534-541.

51. Comtois KA, Chalker SA, Kerbrat AH. Pre- versus postenlistment timing of first suicide attempt as a predictor of suicide risk factors in an active duty military population with suicidal thoughts. Military Behavioral Health. 2016;0(0):1-12.

52. Blackmore ER, Munce S, Weller I, Zagorski B, Stansfeld SA, Stewart DE, Caine ED, Conwell Y. Psychosocial and clinical correlates of suicidal acts: results from a national population survey. British Journal of Psychiatry. 2008; 192(4): 279–284.

53. Heikkinen M, Aro H, Lönnqvist J. Recent life events, social support and suicide. Acta Psychiatrica Scandinavica Supplementum. 1994; 377: 65–72.

54. Heikkinen ME, Lonngvist JK. Recent life events in elderly suicide: a nationwide study in Finland. International Psychogeriatrics. 1995; 7(2): 287-300.

55. Conwell Y, Rotenberg M, Caine ED. Completed suicide at age 50 and over. Journal of the American Geriatrics Society. 1990; 38(6): 640-646.

56. Qin P, Mortensen PB. The impact of parental status on the risk of completed suicide. Archives of General Psychiatry. 2003; 60(8): 797-802.

57. Kposowa AJ. Marital status and suicide in the National Longitudinal Mortality Study. Journal of Epidemiology and Community Health. 2000; 54(4): 254–261.

58. Motto JA, Bostrom A. Empirical indicators of near-term suicide risk. Crisis. 1990; 11(1): 52–59.

59. Wyder M, Ward P, De Leo D. Separation as a suicide risk factor. Journal of Affective Disorders. 2009; 116: 208–213.

60. You S, Van Orden KA, Conner KR. Social connections and suicidal thoughts and behavior. Psychology of Addictive Behaviors. 2011. 25(1): 180-184.

61. Bagge CL, Glenn CR, Lee HJ. Quantifying the impact of recent negative life events on suicide attempts. Journal of Abnormal Psychology. 2013;122(2):359-368.

62. Hooley JM, Franklin JC, Nock MK. Chronic pain and suicide: understanding the association. Current Pain and Headache Reports. 2014; 18:435.

63. Wilson KG, Kowal J, Henderson PR, McWilliams LA, Péloquin K. Chronic pain and the interpersonal theory of suicide. Rehabilitation Psychology. 2013; 58(1): 111-115.

64. Ratcliffe GE, Enns MW, Belik SL, Sareen J. Chronic pain conditions and suicidal ideation and suicide attempts: an epidemiologic perspective. Clinical Journal of Pain. 2008; 24(3): 204-210.

65. Harwood DM, Hawton K, Hope T, Harriss L, Jacoby R. Life problems and physical illness as risk factors for suicide in older people: a descriptive and case-control study. Psychological Medicine. 2006; 36(9): 1265–1274.

66. Fergusson DM, Woodward LJ, Horwood LJ. Risk factors and life processes associated with the onset of suicidal behaviour during adolescence and earl adulthood. Psychological Medicine. 2000;30:23-39.

67. Molnar BE, Berkman LF, Buka SL. Psychopathology, child sexual abuse and other childhood adversities: Relative links to subsequent suicidal behavior in the US. Psychological Medicine. 2001;31:965-977.

68. Nelson EC, Heath AC, Madden PAF, Cooper ML, Dinwiddie SH, Bucholz KK, Glowinski A, McLaughlin T, Dunne MP, Statham DJ, Martin NG. Association between self-reported child sexual abuse and adverse psychosocial outcomes: Results from a twin study. Archives of General Psychiatry. 2002;59:139-145.

69. Belik SL, Stein MB, Asmundson GJ, Sareen J. Relation between traumatic events and suicide attempts in Canadian military personnel. Canadian Journal of Psychiatry. 2009; 54(2): 93-104.

70. Nock MK, Kessler RC. Prevalence of and risk factors for suicide attempts versus suicide gestures: analysis of the National Comorbidity Survey. Journal of Abnormal Psychology. 2006; 115(3):616-623.

71. Brodsky BS, Stanley B. Adverse childhood experiences and suicidal behavior. Psychiatric Clinics of North America. 2008; 31(2), 223-235.

72. Bruffaerts R, Demyttenaere K, Borges G, Haro JM, Chiu WT, Hwang I, Karam EG, Kessler RC, Sampson N, Alonso J, Andrade LH, Angermeyer M, Benjet C, Bromet E, de Girolamo G, de Graaf R, Florescu S, Gureje O, Horiguchi I, Hu C, Kovess V, Levinson D, Posada-Villa J, Sagar R, Scott K, Tsang A, Vassilev SM, Williams DR, Nock MK. Childhood adversities as risk factors for onset and persistence of suicidal behaviour. British Journal of Psychiatry. 2010; 197(1), 20-27.

73. Cox DW, Ghahramanlou-Holloway M, Szeto EH, Greene FN, Engel C, Wynn GH, Bradley J, Grammer G. Gender differences on documented trauma histories: inpatients admitted to a military psychiatric unit for suicide-related thoughts or behaviors. Journal of Nervous and Mental Disease. 2011; 199(3): 183-190.

74. Afifi TO, Boman J, Fleisher W, Sareen J. The relationship between child abuse, parental divorce, and lifetime mental disorders and suicidality in a nationally representative adult sample. Child Abuse and Neglect. 2009; 33(3): 139-147.

75. Afifi TO, MacMillan HL, Boyle M, Taillieu T, Cheung K, Sareen J. Child abuse and mental disorders in Canada. CMAJ. 2014; 186(9): E324-332.

76. Afifi TO, Tailleu T, Zamorski MA, Turner S, Cheung K, Sareen J. Association of child abuse exposure with suicidal ideation, suicide plans, and suicide attempts in military personnel and the general population in Canada. JAMA Psychiatry. 2016;73(3):229-238.

77. Jeon HJ, Park JI, Fava M, Mischoulon D, Sohn JH, Seong S, Park JE, Yoo I, Cho MJ. Feelings of worthlessness, traumatic experience, and their comorbidity in relation to lifetime suicide

attempt in community adults with major depressive disorder. Journal of Affective Disorders. 2014; 166:206-212.

78. O'Brien BS, Sher L. Military sexual trauma as a determinant in the development of mental and physical illness in male and female veterans. International Journal of Adolescent Medicine and Health. 2013; 25(3): 269-274.

79. Sareen J, Cox BJ, Afifi TO, Stein MB, Belik SL, Meadows G, Asmundson GJG. Combat and peacekeeping operations in relation to prevalence of mental disorders and perceived need for mental health care. Archives of General Psychiatry. 2007;64(7):843-852.

80. Bedi S, Nelson EC, Lynskey MT, McCutcheon VV, Heath AC, Madded AF, Martin NG. Risk for suicidal thoughts and behavior after childhood sexual abuse in women and men. Suicide and Life-Threatening Behavior. 2011;41(4):406-415.

81. Araújo RMF, Lara DR. More than words: the association of childhood emotional abuse and suicidal behaviour. European Psychiatry. 2016;37:14-21.

82. Martin MS, Dykxhoorn J, Afifi TO, Colman I. Child abuse and the prevalence of suicide attempts among those reporting suicide ideation. Social psychiatry and psychiatric epidemiology. 2016;51(11):1477-84.

83. Stein DJ, Chiu WT, Hwang I, Kessler RC, Sampson N, Alonso J, Borges G, Bromet E, Bruffaerts R, de Girolamo G, Florescu S, Gureje O, He Y, Kovess-Masfety V, Levinson D, Matschinger H, Mneimneh Z, Nakamura Y, Ormel J, Posada-Villa J, Sagar R, Scott KM, Tomov T, Viana MC, Williams DR, Nock MK. Cross-national analysis of the associations between traumatic events and suicidal behavior: Findings from the WHO World Mental Health Surveys. PLoS One. 2010; 5(5): e10574.

84. Koivumaa-Honkanen H, Honkanen R, Viinamaki H, Heikkila K, Kaprio J, Koskenvuo M. Life satisfaction and suicide: a 20-year follow-up study. American Journal of Psychiatry. 2001: 158: 433-439.

85. Waern M, Rubenowitz E, Wilhelmson K. Predictors of suicide in the old elderly. Gerontology. 2003: 49(5): 328-334.

86. Duberstein PR, Conwell Y, Conner KR, Eberly S, Evinger JS, Caine ED. Poor social integration and suicide: Fact or artifact? A case-control study. Psychological Medicine. 2004; 34: 1331–1337.

87. Dervic K, Oquendo MA, Grunebaum MF, Ellis S, Burke AK, Mann JJ. Religiuos affiliation and suicide attempt. American Journal of Psychiatry. 2004; 161:2303–2308.

88. Cantor CH, Slater PJ. Marital breakdown, parenthood, and suicide. Journal of Family Studies. 1995; 1(2): 91-102.

89. Hoyer G, Lund E. Suicide among women related to number of children in marriage. Archives of General Psychiatry. 1993; 50(2): 134-137.

90. Nelson C, Cyr KS, Corbett B, Hurley E, Gifford S, Elhai JD, Richardson JD. Predictors of posttraumatic stress disorder, depression, and suicidal ideation among Canadian Forces personnel in a National Canadian Military Health Survey. Journal of Psychiatric Research. 2011; 45(11): 1483-1488.

91. Schoenbaum M, Kessler RC, Gilman SE, Colpe LJ, Heeringa SG, Stein MB, Ursano RJ, Cox KL. Predictors of suicide and accident death in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS) – Results from the Army Study to Assess Risk and Reslience in Servicemembers (Army STARRS). JAMA Psychiatry. 2014; 71(5):493-503.

92. Public Health Agency of Canada. The human face of mental health and mental illness in Canada 2006. Ottawa (ON): The Agency; 2006. Available at: www.phac-aspc.gc.ca/publicat/human-humain06/index-eng.php.

93. Anestis MD, Bryan CJ. Means and capacity for suicidal behavior: a comparison of the ratio of suicide attempts and deaths by suicide in the US military and general population. Journal of Affective Disorders. 2013;148:42-47.

94. Durkheim E. LE suicide: Etude de sociologie. Paris : F. Alcan; 1897.

95. Baumeister RF. Suicide as escape from self. Psychological Review. 1990; 97(1): 90-113.

96. Shneidman ES. Perspectives on suicidology: Further reflections on suicide and psychache. Suicide and Life Threatening Behavior. 1998; 28(3): 245-250.

97. Beck AT, Steer RA, Kovacs M, Garrison B. Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. American Journal of Psychiatry. 1985;142(5):559-563.

98. Beck AT, Brown G, Berchick RJ, Stewart BL, Steer RA. Relationship between hopelessness and ultimate suicide: a replication with psychiatric outpatients. American Journal of Psychiatry. 1990; 147(2): 190-195.

99. Linehan MM. Cognitive-behavioral treatment of borderline personality disorder. New York: Guilford; 1993.

100. Joiner T. Why people die by suicide. Cambridge (MA): Harvard University Press; 2005.

101. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Xieyining H, Musacchio KM, Jaroszewski AC, Chang BP, Nock MK. Risk foators for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychological Bulletin. 2017;143(2):187-232.

102. Klonsky ED, May AM. Differentiating suicide attempters from suicide ideators: a critical frontier for suicidology research. Suicide and Life-Threatening Behavior. 2014;44(1):1-5.

103. Selby E, Joiner TE, Ribeiro JD. Comprehensive theories of suicidal behaviors. The Oxford Handbook of Suicide and Self-Injury. 2014;286.

104. Stanley IH, Hom MA, Rogers ML, Hagan CR, Joiner TE. Understanding suicide among older adults: a review of psychological and sociological theories of suicide. Aging & Mental Health. 2016;20(2):113-122.

105. McMillan D, Gilbody S, Beresford E, Neilly L. Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. Psychological Medicine. 2007;37(6):769-778.

106. Van Orden KA, Witte TK, Gordon KH, Bender TW, Joiner TE. Suicidal desire and the capability for suicide: tests of the interpersonal-psychological theory of suicidal behaviour among adults. Journal of Consulting and Clinical Psychology. 2008; 76(1): 72-83.

107. Joiner TE, Van Orden KA, Witte TK, Selby EA, Ribeiro JD, Lewis R, Rudd D. Main predictions of the interpersonal-psychological theory of suicidal behavior: empirical tests in two samples of young adults. Journal of Abnormal Psychology. 2009; 118(3): 634-646.

108. Christensen H, Batterham PJ, Soubelet A, Mackinnon AJ. A test of the interpersonal theory of suicide in a large community-based cohort. Journal of Affective Disorders. 2013; 144(3): 225-234.

109. Christensen H, Batterham PJ, Mackinnon AJ, Donker T, Soubelet A. Predictors of the risk factors for suicide identified by the interpersonal-psychological theory of suicidal behavior. Psychiatry Research. 2014;219:290-297.

110. Cukrowicz KC, Cheavens JS, Van Orden KA, Ragain RM, Cook RL. Perceived burdensomeness and suicide ideation in older adults. Psychology and Aging. 2011. 26(2) : 331-338.

111. Anestis MD, Bagge CL, Tull MT, Joiner TE. Clarifying the role of emotion dysregulation in the interpersonal-psychological theory of suicidal behavior in an undergraduate sample. Journal of Psychiatric Research. 2011;45:603-611.

112. Barzilay S, Feldman D, Snir A, Apter A, Carli V, Hoven CW, Wasserman C, Sarchiapone N, Wasserman D. The interpersonal theory of suicide and adolescent suicidal behavior. Journal of Affective Disorders. 2015;183:68-74.

113. Bryan CJ, Morrow CE, Anestis MD, Joiner TE. A preliminary test of the interpersonal-psychological theory of suicidal behavior in a military sample. Personality and Individual Differences. 2010a;48:347-350.

114. Bryan CJ, Cukrowicz KC, West CL, Morrow CE. Combat experience and acquired capability of suicide. Journal of Clinical Psichology. 2010b;66(10):1044-1056.

115. Bryan CJ, Cukrowiza KC. Associations between types of combat violence and the acquired capability for suicide. Suicide and Life-Threatening Behavior. 2011a;41(2):126-136.

116. Bryan CJ, Anestis M. Reexperiencing symptoms and the interpersonal-psychological theory of suicidal behavior among deployed service members evaluated for traumatic brain injury. Journal of Clinical Psychology. 2011b;67(9):856-865.

117. Bryan CJ. The clinical utility of a brief measure of perceived burdensomeness and thwarted belongingness for the detection of suicidal military personnel. Journal of Clinical Psychology. 2011c;67(10):981-992.

118. Bryan CJ, Clemans TA, Hernandez AM. Perceived burdensomeness, fearlessness of death, and suicidality among deployed military personnel. Personality and Individual Differences. 2012;52: 374-379.

119. Anestis MD, Khazem LR, Mohn RS, Green BA. Testing the main hypotheses of the interpersonal-psychological theory of suicidal behavior in a large diverse sample of United States military personnel. Comprehensive Psychiatry. 2015;60:78-85.

120. Van Orden KA, Lynam ME, Hollar D, Joinet TE. Perceived burdensomeness as an indicator of suicidal symptoms. Cognitive Therapy Research. 2006; 30: 457-467.

121. Van Orden KA, Witte TK, Gordon KH, Bender TW, Joiner TE. Suicidal desire and the capability for suicide: tests of the interpersonal-psychological theory of suicidal behaviour among adults. Journal of Consulting and Clinical Psychology. 2008; 76(1): 72-83.

122. Statistics Canada. Canadian Forces Mental Health Survey (CFMHS). 2013. Available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5084.

123. Statistics Canada. Canadian Community Health Survey – Mental Health (CCHS). 2013. Available at: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5015.

124. Zamorski MA. Suicide prevention in military organizations. International Review of Psychiatry. 2011; 23(2): 173-180.

125. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry. 2016;3(7):646-659.

126. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E et al. Evidence-based national suicide prevention taskforce in Europe: a consensus position paper. European Neuropsychopharmacology. 2017;27(4):418-421.

127. Sareen J, Afifi TO, Taillieu T, Cheung K, Turner S, Bolton SL, Erickson J, Stein MB, Fikretoglu D, Zamorski MA. Trends in suicidal behaviour and use of mental health services in Canadian military and civilian populations. CMAJ. 2016; 188(11):E261-E267.

128. Kessler RC, Berglund P, Borges G, Nock M, Wang PS. Trends in suicide ideation, plans, gestures, and attempts in the United States, 1990-1992 to 2001-2003. JAMA. 2005;293(20):2487-2495.

129. Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, Nock MK. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychological Medicine. 2016;46:225-236.

130. World Health Organization. Public health action for prevention of suicide: a framework. Geneva, Switzerland; 2012.

131. Nock MK. Recent and needed advances in the understanding, prediction, and prevention of suicidal behavior. Depression and anxiety. 2016;33(6):460-463.

132. Mann JJ, Ellis SP, Waternaux CM, Liu X, Oquendo MA, Malone KM, Brodsky BS, Haas GL, Currier D. Classification trees distinguish suicide attempters in major psychiatric disorders: a model of clinical decision making. Journal of Clinical Psychiatry. 2008;69(1):23-31.

133. Lopez-Castroman J, Perez-Rodriguez MM, Jaussent I, Alegria AA, Artes-Rodriguez A et al. Distinguishing the relevant features of frequent suicide attempters. Journal of Psychiatric Research. 2011;45(5):619-625.

134. Delgado-Gomez D, Blasco-Fontecilla H, Sukno F, Ramos-Plasencia MS, Baca-Garcia E. Suicide attempters classification: toward predictive models of suicidal behavior. Neurocomputing. 2012;92:3-8.

135. Kessler RC, Warner CH, Ivany C, Petukhova MV, Rose S et al. Predicting suicides after psychiatric hospitalization in US Army soldiers – The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). JAMA Psychiatry. 2015;72(1):49-57.

136. Kessler RC, Stein MB, Petukhova MV, Bliese P, Bossarte RM et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Molecular Psychiatry. 2016;00:1-8.

137. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science 2017;5(3):457-469.

138. Chang BP, Franklin JC, Ribeiro JD, Bentley KH, Kleiman EM, Nock MK. Biological risk factors for suicidal behaviors: a meta-analysis. Translational Psychiatry. 2016;6:887.

139. Bentley KH, Franklin JC, Ribeiro JD, Kleiman EM, Fox KR, Nock MK. Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: a meta-analytic review. Clinical Psychology Review. 2016;43:30-46.

140. Clark LA, Watson D. Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. Journal of Abnormal Psychology. 1991;100(3):316-336.

141. Feldman LA. Distinguishing depression and anxiety in self-report: Evidence from confirmatory factor analysis on nonclinical and clinical samples. Journal of Consulting and Clinical Psychology. 1993;61(4):631-638.

142. Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. Annual Review of Psychology. 1998;49:377-412.

143. McLaughlin R, Nielsen L, Waller M. An evaluation of the effect of military service on mortality: quantifying the health soldier effect. Annals of Epidemiology. 2008: 18(12): 928-936.

144. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of Clinical Epidemiology. 2015;68:25-34.

145. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.

146. Azen R, Budescu DV. The dominance analysis approach for comparing predictors in multiple regression. Psychological Methods. 2003;8(2):129-148.

147. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis. 2008;52:2249-2260.

148. Rokach, L, Maimon O. Data mining with decision trees: theory and applications. World Scientific Publishing; 2008.

149. Louppe G. Understanding random forests: from theory to practice. PhD thesis, University of Liege; 2014.

150. Japkowicz N, Shah M. Evaluating learning algorithms: a classification perspective. Cambridge University Press; 2011.

151. Hastie TJ, Tibshirani RJ, Friedman JJH. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. Springer-Verlag New York; 2009.

152. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. New York: Springer; 2013.Japkowicz N, Stephen S. The class imbalance problem: a systematic study. Intelligent Data Analysis. 2002;6(5): 429-449.

153. Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013.

154. Boulesteix AL, Janitza S, Kruppa J, König I. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2012;2(6):493-507.

155. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27:861-874.

156. Malley JD, Malley KG, Pajevic S. Statistical learning for biomedical data. Cambridge University Press; 2011.

157. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intelligent Data Analysis. 2002;6:429-449.

158. Blagus R, Lusa L. Class prediction for high-dimensional class-imbalanced data. BMC Bioinformatics. 2010;11:523.

159. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002;16:321-357.

160. Hulse JV, Khoshgoftaar TM, Napolitano A. Experimental perspectives on learning from imbalanced data. Proceedings of the 24th International Conference on Machine Learning. Oregon State University, 2007.

161. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics. 2013;14:106.

162. Nadeau C, Bengio Y. Inference for the generalization error. Machine Learning. 2003;52:239–281.

163. Bouckaert RR. Choosing between two learning algorithms based on calibrated tests. Proceedings of 20th International Conference on Machine Learning. Morgan Kaufmann, 2003.

164. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics. 2006;7: 3.

165. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recognition Letters. 2010;31(14):2225–2236.

166. Altmann A, Tolosi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010. 26(10):1340–1347.

167. Daelemans W, Hoste V, De Meulder F, Naudts B. Combined optimization of feature selection and algorithm parameters in machine learning of language. Proceedings of the 14th European Conference on Machine Learning, 2003.

168. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological methods. 2009;14(4):323.

169. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat. 2006;15(3): -674.

170. Strobl C. Data mining. The Oxford Handbook of Quantitative Methods in Psychology, Vol. 2: Statistical Analysis. Oxford University Press, 2015.

171. Loh WY, Shih YS. Split selection methods for classification trees. Statistica Sinica. 1997;7:815-840.

172. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics. 2007;8:25.

173. Kim H, Loh W. Classification trees with unbiased multiway splits. Journal of the American Statistical Association. 2001;96, 589–604.

174. Breiman L. Bagging predictors. Machine learning. 1996;24(2):123-140.

175. Breiman L. Random forests. Machine learning. 2001;45(1):5–32.

176. Breiman L. Arcing classifiers. The Annals of Statistics. 1998;26(3):801-849.

177. Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning. 1999;36(1-2):105–139.

178. Bühlmann P, Yu B. Analyzing bagging. Annals of Statistics. 2000;30:927–961.

179. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning. 2000;40(2):139–157.

180. Wu B, Abbott T, Fishman D, McMurray W, Mor G, Stone K, Ward D, Williams K, Zhao H. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics. 2003;19(13):1636–1643.

181. Lee JW, Lee JB, Park M, Song SH. An extensive evaluation of recent classification tools applied to microarray data. Computation Statistics and Data Analysis. 2005;48:869–885.

182. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd International Conference on Machine Learning. ACM Press; 2006.

183. Qi Y, Bar-Joseph Z, Klein-Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins. 2006;63: 490–500.

184. Lunetta KL, Hayward LB, Segal J, Eerdewegh PV. Screening large-scale association study data: exploiting interactions using random forests. BMC Genetics. 2004;5:32.

185. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008;9:307.

186. Nicodemus KK, Malley JD. Predictor correlation impacts machine learning algorithms: implications for genomic studies. Bioinformatics. 2009;25(15):1884-1890.

187. Hapfelmeier A, Ulm K. A new variable selection approach using random forests. Computational Statistics and Data Analysis. 2013;60:50–69. Hapfelmeier A, Ulm K. A new variable selection approach using random forests. Computational Statistics and Data Analysis. 2013;60:50–69.

188. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington (DC): American Psychiatric Press Inc; 2000.

189. Wittchen HU. Reliability and validity studies of the WHO—Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res. 1994;28(1):57–84.

190. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, . Concordance of the Composite International Diagnostic Interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. Int J Methods Psychiatr Res. 2006;15(4):167–180.

191. Meadows G, Burgess P, Fossey E, Harvey C. Perceived need for mental health care, findings from the Australian National Survey of Mental Health and Well-being. Psychological Medicine. 2000;(30); 645–56.

192. Schwartz JE, Pieper CF, Karasek RA. A procedure for linking psychosocial job characteristics data to health surveys. American Journal of Public Health. 1988;78:904–909.

193. Karasek R, Theorell T. Healthy work: Stress, productivity and the reconstruction of working life. Basic Books, 1990.

194. Cutrona, CE, Russell DW. The provisions of social relationships and adaptation to stress. Advances in Pers Relationships. 1987;1(1):37–67.

195. Walsh CA, MacMillan HL, Trocmé N, Jamieson E, Boyle MH. Measurement of victimization in adolescence: Development and validation of the Childhood Experiences of Violence Questionnaire. Child abuse & neglect. 2008;32(11), 1037-1057.

196. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R version 3.3.2. Available at https://www.R-project.org/.

197. Kuhn M. Building predictive models in R using the caret package. Journal of Statistical Software 2008; 28(5): 1-26.

198. Kuhn M. caret: Classification and regression training. R package version 6.0–70. Available at http://CRAN.R-project.org/package=caret.

199. Hothorn T, Hornik K, Zeileis A. party: a laboratory for recursive part(y)itioning. 2016. [R package version 1.1–0]. http://CRAN.R-project.org/.

200. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. Available at http://CRAN.R-project.org/package=pROC.

201. Nicodemus KK, Malley J, Strobl C, Ziegler A. The behavior of random forest permutation-based variable importance measures under predictor correlation. BMC Bioinformatics. 2010, 11: 110-10.

202. Genuer R, Poggi JM, Tuleau-Malot C. VSURF: An R package for variable selection using random forests. The R Journal. 2015;7(2):19–33.

203. Strobl C, Hothorn T, Zeileis A. Party on! A new, conditional variable importance measure available in the party package. The R Journal. 2009;2:14-17.

204. Glaser B, Nikolov I, Chubb D, Hamshere ML, Segurado R, Moskvina V, Holmans P. Analyses of single marker and pairwise effects of candidate loci for rheumatoid arthritis using logistic regression and random forests. BMC proceedings. 2007;1(1): S54.

205. Genuer R, Poggi JM, Tuleau C. Random Forests: some methodological insights. arXiv preprint arXiv:0811.3619; 2008.

206. Goldstein BA, Hubbard AE, Cutler A, Barcellos LF. An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. BMC genetics. 2010;11(1):49.

207. Goldstein BA, Polley EC, Briggs F. Random forests for genetic association studies. Statistical applications in genetics and molecular biology. 2011;10(1).

208. Gregorutti B, Michel B, Saint-Pierre P. Correlation and variable importance in random forests. Statistics and Computing. 2017;27(3):659-78.

209. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of Internal Medicine. 1999;130:515-524.

210. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDK. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of Clinical Epidemiology. 2001;54:774-781.

211. Baca-Garcia E, Perez-Rodriguez MM, Oquendo MA, Keyes KM, Hasin DS, Grant BF, Blanco C. Estimating risk for suicide attempt: are we asking the right questions? Passive suicidal ideation as a marker for suicidal behavior. Journal of Affective Disorders. 2011;134(1-3):327-332.

212. Brown GK, Beck AT, Steer RA, Grisham JR. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. Journal of Consulting and Clinical Psychology. 2000;68(3):371-377.

213. Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P,Wells KB, Zaslavsky AM. Prevalence and treatment of Mental Disorders, 1990 to 2003. New England Journal of Medicine. 2005;352:2515-23.

214. Elhai JD, Ford JD. Correlates of mental health service use intensity in the National Comorbidity Survey and National Comorbidity Survey Replication. Psychiatric Services. 2007;58:1108-1115.

215. Merikangas KR, He J, Burstein ME, Swendsen J, Avenevoli S, Case B, Georgiades K, Heaton L, Swanson S, Olfson M. Service utilization for lifetime mental disorders in U.S. adelescents: Results of the National Comorbidity Survey Adolescent Supplement (NCS-A). Journal of the American Academy of Child and Adolescent Psychiatry. 2011;50(1):32-45.

216. Kessler RC, Merikangas KR, Berglund P, Eaton WW, Koretz DS, Walters EE. Mild disorders should not be eliminated from the DSM-V. Archives of General Psychiatry. 2003;60(11):1117-1122.

217. Zamorski MA, Bennett RE, Boulos D, Garber BG, Jetly R, Sareen J. The 2013 Canadian Forces Mental Health Survey: Background and methods. The Canadian Journal of Psychiatry. 2016;61(S1):10S-25S.

218. Fikretoglu D, Liu A, Zamorski MA, Jetly R. Perceived need for care and perceived sufficiency of mental health care in the Canadian Armed Forces: changes in the past decade and comparisons to the general population. Canadian Journal of Psychiatry. 2016:61(S1):36S-45S.

219. Narrow WE, Rae DS, Robins LN, Regier DA. Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. Archives of general psychiatry. 2002;59(2):115-123.

220. Kertz SJ, Bigda-Peyton JS, Rosmarin DH, Björgvinsson T. The importance of worry across diagnostic presentations: prevalence, severity and associated symptoms in a partial hospital setting. Journal of Anxiety Disorders. 2012;26:126-133.

221. LeardMann CA, Powell TM, Smith TC, Bell MR, Smith B, Boyko EJ, Hooper TI, Gackstetter GD, Ghamsary M, Hoge CW. Risk factors associated with suicide in current and former US military personnel. JAMA. 2013;310(5):496-506.

222. Cigrang JA, Balderrama-Durbin C, Snyder DK, Talcott GW, Tatum J, Baker M, Cassidy D, Sonnek S, Smith Slep AM, Heyman RE. Predictors of suicidal ideation across deployment: a prospective study. Journal of Clinical Psychology. 2015;71(9):828-842.

223. Ali S, Nathani M, Jabeen S, Yazdani I, Mouton CD, Bailey RK, Mahr M, Pate RJ, Riley WJ. Alcohol: The lubricant to suicidality. Innovations in Clinical Neuroscience. 2013;10(1):20-29.

224. Fergusson DM, Woodward LJ, Horwood LJ. Risk factors and life processes associated with the onset of suicidal behaviour during adolescence and earl adulthood. Psychological Medicine. 2000;30:23-39.

225. Molnar BE, Berkman LF, Buka SL. Psychopathology, child sexual abuse and other childhood adversities: Relative links to subsequent suicidal behavior in the US. Psychological Medicine. 2001;31:965-977.

226. Afifi TO, Enns MW, Cox BJ. Population attributable fractions of psychiatric disorders and suicide ideation and attmepts associated with adverse childhood experiences. American Journal of Public Health. 2008;98:946-952.

227. Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, Nock MK, Smoller JW, Reis BY. Predicting suicidal behaviour from longitudinal electronic health records. American Journal of Psychiatry. 2017;174:154-162.

228. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu MJ, Zhang XY, Zunta-Soares GB, Quevedo J, Kauer-Sant'Anna M, Kapczinski F, Soares JC. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. Journal of Affective Disorders. 2016;193:109-116.

229. Oh J, Yun K, Hwang JH, Chae JH. Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. Fromtiers in Psychiatry. 2017;8:192.

230. Ma J, Batterham PJ, Calear AL, Han J. A systematic review of the predictions of the Interpersonal–Psychological Theory of Suicidal Behavior. Clinical Psychology Review. 2016;46:34-45.

231. Bryan CJ, Ray-Sannerud BN, Morrow CE, Etienne N. Optimism reduces suicidal ideation and weakens the effect of hopelessness among military personnel. Cognitive Therapy and Research. 2013;37:996-1003.

232. Konig IR, Malley JD, Weimar C, Diener HC, Ziegler A. Practical experiences on the necessity of external validation. Statistics in Medicine. 2007;26:5499-5511.

233. Toth D, Eltinge JL. Building consistent regression trees from complex sample data. Journal of the American Statistical Association. 2011;106(496):1626-1636.

9. Appendix

Table A1. Description of commonly assessed covariates† in CCHS-MH and CFMHS surveys

Index	Description	Parent variable(s)	Values‡
SOCIO-DEMOGRAPHICS			
1	Living arrangement	DHHDHSZ	Living with others Living alone
2	Sex	DHH_SEX SEX	Male Female
3 4 5 6 7	Age	DHH_AGE	Less than 20 years[1] 20-29 years[1] 30-39 years[1] 40-49 years[1] >=50 years[1]
8	Marital status	DHH_MS MARSTAT	Married Common-law Widowed[2] Separated[2] Divorced[2] Single
9	Sexual orientation	SDC_7AA SDC_14	Heterosexual Homosexual[2] Bisexual[2]
10 11	Cultural/racial background	SDCDCGT	White Non-white[1] Multiple[1]
12	Personal education	EDUDR04	Less than secondary Secondary Some post Post-secondary[1]
13 14 15	Working status	LBSDPFT	Did not work Full-time Part-time[1] Not applicable[1] (age<15 or >74)
16	Difficulty meeting basic expenses with current household income	INC_12	No Yes
17	Personal income	INCDPER	No income[2] <5000[2] 5000-9999[2] 10000-14999[2] 15000-19999[2] 20000-29999[2] 30000-39999 40000-49999 50000-59999 60000-69999 70000-79999 80000-89999 90000-99999 100,000 or more
ALCOHOL USE, ABUSE, AND DEPENDENCE			
18	Past-year alcohol consumption	AUD_02	No alcohol <1/month 1/month

			2-3/month 1/week 2-3 /week 4-6/week Daily[1]
19	5 or more drinks on one occasion during past year	AUD_03	No alcohol Never <1/month 1/month 2-3/month 1/week >1/week[1]
20	Ever consulted health care professional about alcohol problems	AUD_39	Yes No
21	Ever received treatment for alcohol problems	AUD_41	Same as above
22	Ever hospitalized for alcohol problems	AUD_42	Same as above
23	Alcohol interference past 12 months	AUDFINT	No Yes
24	Strong urge to drink (lifetime)	AUD_19A	No Yes
25	Need more to get an effect (lifetime)	AUD_19A1	Same as above
26	Experienced withdrawal symptoms or drink to keep symptoms away	AUD_19B AUD_19C	Same as above
27	Used more alcohol than intended	AUD_19D AUD_19E	Same as above
28	Attempts to quit	AUD_19G AUD_29	Same as above
29	Spent days drinking/recovering	AUD_19H	Same as above
30	Reduced activity	AUD_19I	Same as above
31	Continued drinking despite serious problem	AUD_19J	Same as above
32	Alcohol abuse interfered with responsibilities (lifetime)	AUD_13A	Same as above
33	Alcohol abuse despite causing problems to family/friends/neighbors/co-workers (lifetime)	AUD_13B	Same as above
34	Could have been hurt due to alcohol abuse (lifetime)	AUD_13C	Same as above
35	Number of times stopped by the police	AUD_13E	None Once Multiple times[1]
36 37	Past-year alcohol abuse or dependence	AUDDY AUDDL	No More than one year ago[1] Past-year [1]
	GENERALIZED ANXIETY DISORDER		
38	Worried about everything	GAD_01A	No Yes
39	Worried about nothing in particular	GAD_01B	Same as above
40	Worried about finances	GAD_01C	Same as above
41	Worried about success at school or work	GAD_01D	Same as above
42	Worried about social life	GAD_01E	Same as above
43	Worried about love life	GAD_01F	Same as above
44	Worried about relationships at school	GAD_01G	Same as above

	or work		
45	Worried about relationships with family	GAD_01H	Same as above
46	Worried about physical appearance	GAD_01I	Same as above
47	Worried about own physical health	GAD_01J	Same as above
48	Worried about own mental health	GAD_01K	Same as above
49	Worried about alcohol or drug use	GAD_01L	Same as above
50	Worried about being away from home or apart from loved ones	GAD_01M	Same as above
51	Worried about the health or welfare of loved ones	GAD_01N	Same as above
52	Worried about social phobias	GAD_01O	Same as above
53	Worried about agoraphobia	GAD_01P	Same as above
54	Worried about specific phobias	GAD_01Q	Same as above
55	Worried about obsessions	GAD_01R	Same as above
56	Worried about compulsions	GAD_01S	Same as above
57	Worried about crime/violence	GAD_01T	Same as above
58	Worried about economy	GAD_01U	Same as above
59	Worried about environment	GAD_01V	Same as above
60	Worried about moral decline of society	GAD_01W	Same as above
61	Worried about war/revolution	GAD_01X	Same as above
62	Worried about other	GAD_01Y	Same as above
63	Difficulty to control anxiety feelings	GAD_04	Often[1] Sometimes Rarely Never
64	Clinically significant distress - patient so nervous cannot think at anything else	GAD_04A	Often[1] Sometimes Rarely Never
65	Emotional distress intensity (lifetime)	GAD_15	None Mild Moderate Severe Very severe[1]
66	Degree of interference with work/social life/personal relationships (lifetime)	GAD_17	Not at all A little Some A lot Extremely[1]
67	Unable to carry out daily activities	GAD_17A	Often[1] Sometimes Rarely Never
68	Past-year interference with home responsibilities	GAD_38A	0 (No interference) 1-9[2] 10 (Very severe interference)[2]
69	Past-year interference with close relationships	GAD_38C	Same as above
70	Past-year interference with social life	GAD_38D	Same as above
71	6 months or more of worry or anxiety	GADDA2	Same as above
72	Anxiety-associated restlessness	GAD_09A	Same as above
73	Anxiety-associated fatigue	GAD_09B	Same as above
74	Anxiety-associated irritability	GAD_09C	Same as above
75	Anxiety-associated difficulty concentrating	GAD_09D	Same as above

76	Anxiety-associated tense or aching muscles	GAD_09E	Same as above
77	Anxiety-associated sleep problems	GAD_09F	Same as above
78	Anxiety-associated heart pounding	GAD_10A	Same as above
79	Anxiety-associated sweating	GAD_10B	Same as above
80	Anxiety-associated trembles/shakes	GAD_10C	Same as above
81	Anxiety-associated dry mouth	GAD_10D	Same as above
82	Anxiety-associated sadness/depressed mood	GAD_10E	Same as above
83	Contact with health care professional, treatment and hospitalization due to anxiety	GAD_44 GAD_58 GAD_59	No contact Treatment more than 12 month ago Past-year treatment without hospitalization Past-year treatment and hospitalization[1]
84 85 86 87 88	Number and type of anxiety episodes lasting 6 months or longer	GAD_311 GAD_26D GAD_26C	No episode Single episode out of blue[1] Multiple episodes, none brought on by stress[1] Single episode brought on by stress[1] Multiple episodes with single episode brought on by stress[1] Multiple episodes with multiple episodes brought on by stress[1]
89 90	Type of most recent episode	GAD_27C GAD_27 DHH_AGE	Current episode Recent (1 to 5 years) episode[1] Remote (>5 years ago) episode[1]
91 92	Past-year social or occupational interference	GADD384 GAD_38A GAD_38C GAD_38D	None Sub-threshold interference[1] Threshold interference[1]
93 94	Number of days totally unable during past year	GAD_40	None 1 to 4 days[1] More than 5 days[1]
95 96	Type of past-year interference	GADD384 GAD_38A GAD_38C GAD_38D GAD_40 GAD_58	No interference Sub-threshold interference[1] Threshold interference[1]
97 98 99 100	Past-year GAD	GADDYA GADDYB GAD_27C GAD_27 DHH_AGE GAD_38A GAD_38C GAD_38D GAD_40 GAD_58	No GAD GAD with last episode more than 5 years ago[1] GAD with last episode between 1 to 5 years ago OR with episode within 12m without interference[1] Sub-threshold past-year GAD[1] Past-year GAD[1]
GENERAL HEALTH			
101	Self-perceived health compared to 1 year ago	GEN_02	Much better Somehow better Same Somewhat worse Much worse[1]
102	Trouble sleeping	GEN_04	No A little Some

			Most All the time[1]
103	Perceived life stress	GEN_07	Not at all Not very A bit Quite a bit Extremely[1]
104	Self-perceived work stress	GEN_09	Not at all Not very A bit Quite a bit Extremely[1]
105	Sense of belonging to local community	GEN_10	Very strong Somewhat strong Somewhat weak Very weak[1]
106	Perceived health	GENDHDI	Poor[1] Fair Good Very good Excellent
107	Satisfaction with life in general	GENGSWL	Very satisfied Satisfied Neither Dissatisfied Very dissatisfied[1]
CHRONIC CONDITIONS			
108	Self-reported asthma	CCC_031	No Yes
109	Self-reported arthritis	CCC_051	Same as above
110	Self-reported back problems excluding fibromyalgia and arthritis	CCC_061	Same as above
111	Self-reported migraine headaches	CCC_081	Same as above
112	Self-reported diabetes	CCC_101	Same as above
113	Self-reported post-traumatic stress disorder	CCC_311	Same as above
114	Self-reported other long-term physical or mental health condition	CCC_901	Same as above
115	Self-reported current or past high blood pressure	CCCDHBP	Same as above
116	Self-reported phobia	CCC_29AA CCC290BA	Same as above
117	Self-reported obsessive-compulsive disorder	CCC_29AB CCC290BB	Same as above
118	Self-reported panic disorder	CCC_29AC CCC290BC	Same as above
119	Self-reported other anxiety disorder	CCC_29AD CCC290BD	Same as above
120	Self-reported mood disorder excluding depression (bipolar disorder or mania or dysthymia or other)	CCC_28AA CCC_28AB CCC_28AD CCC_28AE CCC280BA CCC280BB CCC280BC CCC280BD	Same as above

MEDICATION USE			
121	Past 12m medication for emotions/mental health/alcohol use/drug use	MED_01	No Yes
PERCEIVED NEED FOR CARE			
138 139	Past-year perceived need for mental health care	PNCDNEED	No perceived need All perceived needs met Perceived needs partially met[1] Perceived needs not met[1]
122 123 124 125	Received as much help as needed in 3 domains (information, medication, and counseling/therapy)	PNC_02A PNC_01A PNC_01B, PNC_01C	No help (did not need help) Received as much help as needed in all 3 domains[1] Received as much help as needed in 1 or 2 domains[1] Received help in only 1 domain and needed more[1] Received help in 2 or more domains and needed more[1]
PERCEIVED NEED FOR HELP AND HELP NEEDED[5]			
126	Received partial help or no help at all because self-managed	PN1_01A1 to PN1_01A7	No Yes
127	Not aware of help being available	PN1_01B1 to PN1_01B7	Same as above
128	Too busy	PN1_01C1 to PN1_01C7	Same as above
129	Job interfered	PN1_01D1 to PN1_01D7	Same as above
130	Help not readily available	PN1_01E1 to PN1_01E7	Same as above
131	No confidence into system	PN1_01F1 to PN1_01F7	Same as above
132	Help unaffordable	PN1_01G1 to PN1_01G7	Same as above
133	Insurance not cover	PN1_01H1 to PN1_01H7	Same as above
134	Afraid of others opinions	PN1_01I1 to PN1_01I7	Same as above
135	Language problems	PN1_01J1 to PN1_01J7	Same as above
136	Help is ongoing	PN1_01K1 to PN1_01K7	Same as above
137	Other	PN1_01L1 to PN1_01L7	Same as above
WORK STRESS			
140	Past-year job required learning new things	WST_401	Strongly agree Agree Neither agree not disagree Disagree[2] Strongly disagree[2]
141	Past-year job required high level of skill	WST_402	Same as above
142	Past-year job allowed for freedom to decide	WST_403	Same as above
143	Past-year job required repetitive tasks	WST_404	Same as above
144	Past-year job was very hectic	WST_405	Same as above

145	Past-year job was free from conflicting demands	WST_406	Same as above
146	Past-year job security was good	WST_407	Same as above
147	Past-year job allowed for own job input	WST_409	Same as above
148	Supervisor was helpful in getting it done	WST_411	Same as above
149	Co-workers were helpful in getting it done	WST_412	Same as above
150	Required lot of physical effort	WST_408	Strongly agree[2] Agree[2] Neither agree not disagree Disagree Strongly disagree
151	Exposed to hostility or conflict with others	WST_410	Same as above
152	Job satisfaction	WST_413	Very satisfied Somewhat satisfied Not too satisfied Not at all satisfied[1]
SOCIAL PROVISION SCALE			
153	Has people to depend on	SPS_01	Strongly agree Agree Disagree[2] Strongly disagree[2]
154	Has people who enjoy the same activities	SPS_02	Same as above
155	Has close relationships	SPS_03	Same as above
156	Has someone to talk to about important decisions	SPS_04	Same as above
157	Has relationships that recognize competence/skill	SPS_05	Same as above
158	Has trustworthy person for advice	SPS_06	Same as above
159	Feel part of a group who shared attitudes/beliefs	SPS_07	Same as above
160	Has strong emotional bond	SPS_08	Same as above
161	Has people who admire talents/abilities	SPS_09	Same as above
162	Has people to count on emergency	SPS_10	Same as above
163	Level of social provisions scale (based on social provisions scale overall score)	SPSDCON	Low (<30)[1] High (30-39) Highest (40)
DEPRESSION			
164 165	Daily duration of two-week (or longer) period of being sad/discouraged/uninterested	DEP_09 DEP_12 DEP_16 DEP_17	None Less than 5 hours[1] 5 hours or more[1]
166 167 168 169	Intensity of emotional distress during two-week period (or longer) of being sad/discouraged/uninterested	DEP_09 DEP_12 DEP_17	No distress Mild[1] Moderate[1] Severe[1] Very severe[1]
170	Experienced severe emotional distress during two-week (or longer) period	DEP_09 DEP_12 DEP_18 DEP_19	No Yes
171	Had depressive-like symptoms during two-week (or longer) period	DEP_21	Same as above

172 173 174 175 176 177 178 179	Worst/last episode history	DEP_21 DEP_22 DHH_AGE DEP_22A DEP_23A DEP_22B1 DEP_22C1 DEP_23B1 DEP_23C1	No depressive-like symptoms Worst episode stopped more than 5 years ago[1] Worst episode stopped 2 to 5 years ago[1] Worst episode stopped during the year before interview[1] Worst episode currently happening[1] Last bad episode stopped more than 5 years ago[1] Last bad episode stopped 2 to 5 years ago[1] Last bad episode stopped during the year before the interview[1] Last bad episode currently happening[1]
180 181	Sad/empty/depressed most of the day during two-week (or longer) period	DEP_21 DEP_24A DEP_24B	No Yes
182 183	Discouraged/hopeless during two-week (or longer) period	DEP_21 DEP_24C DEP_24D	Same as above
184	Uninterested during two-week (or longer) period	DEP_21 DEP_24E	Same as above
185	No fun during two-week (or longer) period	DEP_21 DEP_24F	Same as above
186	Significant weight loss/gain or change in appetite during two-week (or longer) period	DEPDA23	Same as above
187 188 189 190 191	Insomnia/hypersomnia during two-week (or longer) period	DEP_21 DEP_24A DEP_24B DEP_24C DEP_24D DEP_24E DEP_24F DEP_26G DEP_26H DEP_26I	No Hypersomnia[1] No insomnia/hypersomnia and did not sleep less Than usual[1] No insomnia/hypersomnia but slept less than usual[1] Insomnia but did not sleep much less than usual[1] Insomnia and slept much less than usual[1]
192 193	Fatigue/loss of energy during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26J DEP_26K	No A lot more energy than usual[1] Fatigue or loss of energy[1]
194 195 196 197	Restlessness or retardation during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26L DEP_26M DEP_26N DEP_26O	No Retardation noticed by others Retardation not noticed by others Restlessness not noticed by others Restlessness noticed by others
198 199	Diminished ability to think during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26P DEP_26Q	No Slow in thinking than usual Racing thoughts
200	Trouble concentrating during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26R	No Yes

201	Undecided during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26S	Same as above
202	Lost self-confidence during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26T	Same as above
203	Worthlessness during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26U DEP_26V	No Not totally worthless Totally worthless[1]
204	Guilt during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26W	No Yes
205	Irritability during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26X	Same as above
206	Anxious most of the days during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26Y	Same as above
207	Had attacks of fear or panic during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_26Z	Same as above
208	Unable to cope with everyday responsibilities during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_265	Same as above
209	Less talkative during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_267	Same as above
210	Often in tears during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_268	Same as above
211	Death ideation during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_6A	Same as above
212	Passive suicide ideation during two-week (or longer) period	DEP_21 DEP_24A to DEP_24F DEP_6B	Same as above
213 214 215 216 217 218	Type of two-week episode	DEP_31 DEP_38 DEP_8A	No threshold episodes Single threshold episode more than a year ago[1] Multiple threshold episodes more than a year ago[1] Single threshold episode during past year[1] Multiple episodes, including at least one threshold episode during past year[1]

			Past month single threshold episode[1] Multiple episodes, including one past month threshold episode[1]
219 220 221	Past-year MDE including distress	DEPDYA DEPDDY DEPDYB DEPFINT	No MDE MDE more than one year ago[1] Past-year MDE without significant interference[1] Past-year MDE with significant interference[1]
222 223 224	Bereavement	DEP_30 DEP_30A	No threshold episodes Episodes never triggered by bereavement[1] Some episodes triggered by bereavement[1] Episodes always triggered by bereavement[1]
MENTAL HEALTH EXPERIENCES			
225	Work in a mental health treatment program	MHE_01B	No Yes
226	Worked with someone treated for mental health problem	MHE_02B	Same as above
227	Family member received mental health treatment	MHE_03B	Same as above
228	Friend received mental health treatment	MHE_04B	Same as above
229	Respondent received mental health treatment	MHE_05B	Same as above
230	Discrimination due to mental health	MHE_06	Same as above
231	Personally impacted by stigma experiences	MHEDPIS	Same as above
MENTAL HEALTH SERVICES			
232	Hospitalizations for mental health problems	SR1_002	No hospitalization One 2 or more[1]
233 234 235	Duration of hospitalization for mental health problems	SR1_003	No hospitalization 1 day[1] 2 to 6 days[1] 7 days or more[1]
236	Frequency of consultation with psychiatrist during past year	SR1_010	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
237	Duration of average consultation with psychiatrist during past year	SR1_011	No consultation Less than 60 minutes 60 minutes or more[1]
238	Perceived level of help received from psychiatrist during past year	SR1_012	No consultation A lot Some A little[2] Not at all[2]
239 240 241	Currently stopped seeing psychiatrist	SR1_013 SR1_014A to SR1_014L	No consultation Stopped due to single reason[1] Stopped due to multiple reasons[1] Currently seeing psychiatrist[1]
242	Frequency of consultation with family doctor during past year	SR1_020	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly
243	Duration of average consultation with	SR1_021	No consultation

	family doctor during past year		Less than 60 minutes 60 minutes or more
244	Perceived level of help received from family doctor during past year	SR1_022	No consultation A lot Some A little[2] Not at all[2]
245 246 247	Currently stopped seeing family doctor	SR1_023 SR1_024A to SR1_024L	No consultation Stopped due to single reason[1] Stopped due to multiple reasons[1] Currently seeing family doctor[1]
248	Frequency of consultation with psychologist during past year	SR1_030	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
249	Duration of average consultation with psychologist during past year	SR1_031	No consultation Less than 60 minutes 60 minutes or more[1]
250	Perceived level of help received from psychologist during past year	SR1_032	No consultation A lot Some A little[2] Not at all[2]
251 252 253	Currently stopped seeing psychologist	SR1_033 SR1_034A to SR1_034L	No consultation Stopped due to single reason[1] Stopped due to multiple reasons[1] Currently seeing psychiatrist[1]
254	Frequency of consultation with nurse during past year	SR1_040	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
255	Duration of average consultation with nurse during past year	SR1_041	No consultation Less than 60 minutes 60 minutes or more[1]
256	Perceived level of help received from nurse during past year	SR1_042	No consultation A lot Some A little[2] Not at all[2]
257 258 259	Currently stopped seeing nurse	SR1_043 SR1_044A to SR1_044L	No consultation Stopped due to single reason[1] Stopped due to multiple reasons[1] Currently seeing psychiatrist[1]
260	Frequency of consultation with social worker during past year	SR1_050	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
261	Duration of average consultation with social worker during past year	SR1_051	No consultation Less than 60 minutes 60 minutes or more[1]
262	Perceived level of help received from social worker during past year	SR1_052	No consultation A lot Some A little[2] Not at all[2]
263 264	Currently stopped seeing social worker	SR1_053 SR1_054A to	No consultation Stopped due to single reason[1]

265		SR1_054L	Stopped due to multiple reasons[1] Currently seeing psychiatrist[1]
266	Frequency of consultation with family during past year	SR1_060	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
267	Perceived level of help received from family during past year	SR1_062	No consultation A lot Some A little[2] Not at all[2]
268	Frequency of consultation with friend during past year	SR1_70	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
269	Perceived level of help received from friend during past year	SR1_072	No consultation A lot Some A little[2] Not at all[2]
270	Frequency of consultation with co-workers/supervisor during past year	SR1_080	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
271	Perceived level of help received from co-workers/supervisor during past year	SR1_082	No consultation A lot Some A little[2] Not at all[2]
272	Frequency of consultation with others during past year	SR1_100	No consultation Once Multiple visits, less than once a month Multiple visits, at least monthly[1]
273	Perceived level of help from others during past year	SR1_102	No consultation A lot Some A little[2] Not at all[2]
274	Used internet to learn about symptoms during past year	SR1_112A	No Yes
275	Used internet to find out where to get help during past year	SR1_112B	Same as above
276	Used internet to discuss with others during past year	SR1_112C	Same as above
277	Used internet to get online therapy during past year	SR1_112D	Same as above
278	Used internet for other reasons during past year	SR1_112E	Same as above
279	Went to self-help group during past year	SR1_113	Same as above
280	Went to emotional or mental health type self-group during past year	SR1_114A	Same as above
281	Went to alcohol or drug use type self-group during past year	SR1_114B	Same as above
282	Felt prejudice/discrimination because of personal characteristics during past-	SR1_120	Same as above

	year		
283	Felt prejudice/discrimination because of own mental health problems during past-year	SR1_121	Same as above
284	Needed help but received none during past year	SR1_118 SR1_119	Same as above
	ADEVERSE CHILDHOOD EXPERIENCES		
285 286	Child physical abuse	CEX_02 CEX_03 CEX_04	No Physical abuse not involving sexual assault[1] Physical abuse involving sexual assault[1]
287 288	Child sexual abuse	CEX_05 CEX_06	No Sexual abuse not involving sexual assault[1] Sexual abuse involving sexual assault[1]
289	Witnessing domestic violence	CEX_01	No Yes
290	Lifetime contact with child protection services	CEX_07	No Yes
	SUICIDE		
291	Suicide thoughts during two-week (or longer) period[3]	DEP_21 DEP_24A to DEP_24F SUI_01	No Yes
292	Suicide thoughts outside two-week (or longer) period[3]	DEP_21 DEP_24A to DEP_24F SUI_02	Same as above
293	Suicide plans during two-week (or longer) period[4]	DEP_21 DEP_24A to DEP_24F SUI_04A	Same as above
294	Suicide plans outside two-week (or longer) period[4]	DEP_21 DEP_24A to DEP_24F SUI_04B	Same as above

†All binary covariates were included for variable selection; For multi-categorical variables, see footnotes [1,2]
‡Where applicable, each variable included a missing (don't know/refusal/ not stated) category
[1]Category chosen to be converted into a binary variable for variable selection
[2]Categories chosen to be collapsed and converted into a binary variable for variable selection
[3]Used as a covariate only when the outcome was past-year suicide plans or past-year suicide attempts
[4]Used as a covariate only when the outcome was past-year suicide attempts
[5]Reasons for not seeking help after receiving partial help or no help at all in any of the following domains: information, medication, and counseling/therapy.

Table A.2. Comparison between CFMHS and CCHS-MH surveys [122,123]

Item	CFMHS	CCHS-MH
Target population	Full time regular members of the Canadian Forces and reservists who have been deployed in support of the mission in Afghanistan.	Population 15 years of age and over living in the ten provinces, excluding persons living on reserves and other Aboriginal settlements; full-time members of the Canadian Forces and the institutionalized population (about 3% of the target population).
Stakeholders	Statistics Canada Department of National Defense	Statistics Canada Health Canada Public Health Agency of Canada Provincial Health Ministries
Survey design	Cross-sectional	Cross-sectional
Sample design	Stratified random sampling	Three-stage stratified cluster sampling
Sample size	8165 6696 (Regular Force) 1469 (Reserve Force)	25,113
Response rate	Regular Force: 79.8% Reserve Force: 78.7%	69.8%
Blocks assessed in both surveys	General Health (GEN) Screening Section (SCR) Chronic conditions (CCC) Pain and Discomfort (HUP) Positive Mental Health (PMH) Distress (DIS) Depression (DEP) Suicide (SUI) General anxiety disorder (GAD) Alcohol Use and Dependence (AUD) WHO Disability Assessment Schedule 2.0 (DAS) Two-week disability (TWD) Mental Health Services (SR1) Medication Use (MED) Perceived Need for Care (PNC) Help Needed (PN1) Mental Health Experiences (MHE) Social Provision Scale (SPS) Childhood Experiences (CEX) Labour Force (LF2) Work Stress (WST) Income (INC) Socio-demographic characteristics (SDC) Education (EDU)	
Blocks assessed in either survey	Stress (STR) Panic disorder (PAD) Post-Traumatic Stress (PTS) Military Sexual Trauma (MST) Attitudes Towards Mental Health Care (ATC) Absenteeism/Presenteeism (ABS) Deployment Experiences (DEX) Exposure to Mental Health Training (MHT) Time away (TMA)	Age of respondent (AN3) Height and Weight (HWT) Physical Activity – Short (PHS) Stress – Sources (STS) Mania (MIA) Smoking (SMK) Substance Use and Dependence (SUD) Drug Identification Numbers (DIN) Family Mental Health Impact (FMI) Negative Social Interactions (NSI) Contact with Police (CWP) Spirituality (SPI)

Table A3. R-Code samples

```
library(caret)
library(party)
library(pROC)
#-----------------------------------#
db<-read.csv(file='Z:DBs/db.csv')
ind<-createDataPartition(db$outcome,p=.5,list=F,times=1)
db1<-db[ind,]
db2<-db[-ind,]
#-----------------------------------#
db1[]<-lapply(db1,factor)
db2[]<-lapply(db2,factor)
#-----------------------------------#
feature.names=names(db1)
for (f in feature.names){
if (class(db1[[f]])=="factor"){
levels<-unique(c(db1[[f]]))
db1[[f]]<-factor(db1[[f]],
labels=make.names(levels))}}
#-----------------------------------#
feature.names=names(db2)
for (f in feature.names){
if (class(db2[[f]])=="factor"){
levels<-unique(c(db2[[f]]))
db2[[f]]<-factor(db2[[f]],
labels=make.names(levels))}}
#-----------------------------------#
set.seed(2017)
rf1 <- train(as.factor(outcome)~., data=db1, method="cforest", metric="ROC",
tuneGrid= expand.grid(.mtry= floor(sqrt(ncol(db1)))), trControl=
trainControl(method="repeatedcv", repeats=10, sampling="down",
classProbs=TRUE, summaryFunction=twoClassSummary, savePredictions=T),
controls=cforest_unbiased(ntree=500), maximize=T)
print(rf1)
#-----------------------------------#
set.seed(2017)
pred1<-predict(rf1,newdata=db2)
cm1<-confusionMatrix(pred1,db2$outcome,positive="X2")
print(cm1)
#-----------------------------------#
pred2<-as.vector(predict(rf1,newdata=db2,type="prob")[,"X2"])
rocrf1<-pROC::roc(db2$outcome,pred2)
print(rocrf1)
ci(rocrf1)
plot(rocrf1)
#-------------------------------#
```