Understanding the differential impact of vegetation measures on the association between vegetation and mental health disorders

by

Abu Yousuf Md Abdullah

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Public Health and Health Systems

Waterloo, Ontario, Canada, 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of my thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract:

Background: Considerable debate exists as to whether vegetation can help achieve better mental health outcomes. Although few studies have attempted to evaluate the health effects of vegetation, a spatial study, which has analyzed the effect of different vegetation measures on the detection of a significant association between vegetation and mental health disorders, is still missing. Furthermore, based on the available literature, there is an absence of studies that have analyzed the age and sex-specific effects of surrounding vegetation on mental health disorders, while adjusting for the overdispersion, spatial autocorrelation and unmeasured covariates in the models.

Objective: The objective of this study is to understand the differential impact of vegetation measures on the association between vegetation and various types of mental health disorders. In doing so, the study also attempted to understand whether there are any age and sex-specific effects of vegetation on mental health disorder cases.

Methods: Remote sensing and machine learning techniques were employed to generate three vegetation indices and one area-based vegetation measure from the Landsat-8 satellite images. The satellite-based indices comprised of the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and the soil-adjusted vegetation index (SAVI). The area-based vegetation measure was developed from a Land Use/Land Cover (LULC) model using the Random Forest ensemble classifier. The conventionally used vegetation data was extracted from the Toronto Open Data portal and compared with the variables created from the satellite images.

The dataset comprising psychotic, non-psychotic, substance use and family, social and occupational-related disorder cases were retrieved from the Ontario Community Health Profiles Partnership database. The dataset also contained the combined mental health disorder cases, which is a total of the four types of mental health disorders.

The association between vegetation and psychotic and non-psychotic disorders were analyzed using the Poisson lognormal models under a Bayesian framework. Based on the results from the Bayesian models, a single vegetation measure was selected and the association of the vegetation with the combined mental health disorders for males and females in the age groups, 0-19, 20-44, 45-64 and 65+ were analyzed using Bayesian spatial modeling.

Results: Results suggested substantial effects of the type of vegetation measure used to analyze the association between vegetation and mental health disorder cases. Only the vegetation indices, which could capture both the areal extent and health of the vegetation cover, could detect a significant association with the mental health disorder cases. Specifically, EVI and SAVI, which were constructed after adjusting for different urban and environmental disturbances, were able to detect significant and negative associations with the psychotic and non-psychotic disorder cases.

Furthermore, the findings of this study suggested significant age and sex-specific effects of vegetation on the prevalence of mental health disorders in Toronto. The combined mental health disorder cases for males from the age group 0-19 years and for both males and females from the age group 20-44 years were found to be negatively associated with the vegetation cover. For older adults in the age-groups 45-64 and 65+, only the socioeconomic covariates were found to be significantly associated with the combined mental health disorder cases.

For each of the Bayesian models analyzed in this study, a substantial influence of the spatially structured and unmeasured covariates was detected.

Conclusions: Epidemiological studies must consider both the quantity and quality of people's exposure to surrounding vegetation cover. Vegetation measures that capture both the areal extent and the health of the surrounding vegetation can help detect the actual relationship between vegetation and the mental health conditions of the people in an area. The study setting (urban, peri-urban and rural) can have a notable influence on the detection of different types of vegetation cover and should always be addressed while selecting a vegetation measure for epidemiological studies. As significant and negative associations between vegetation and mental health disorder cases were found for young males and females, policymakers should consider incorporating more greenspaces and vegetation-covered areas in urban areas, to reduce the future burden of mental health disorders in Canada. The findings of this study can provide critical guidelines to public health researches aiming to understand the exposure of the population to surrounding greenness. The relative risk maps can help devise targeted intervention strategies to reduce mental health burdens in the Toronto area.

Acknowledgments:

First, I would like to express my sincerest gratitude to Dr. Jane Law for all her guidance and supports throughout my Master's program. Dr. Law believed in my capacity to understand the Bayesian Spatial Modeling techniques and has guided me to comprehend some complex concepts related to chronic disease modeling. Even when I had doubts about myself, Dr. Law was patient and encouraged me to carry out the necessary tasks, which ultimately led to the completion of the thesis and this report. I am truly fortunate to have received Dr. Law's mentorship and firmly believe that her teachings would greatly assist me in achieving my future career goals in academia.

Second, I am truly grateful to my committee members, Dr. Zahid A. Butt and Dr. Chris Perlman. Dr. Butt had guided me to understand the studied variables thoroughly and provided valuable instructions in both the study design and the interpretation of the results. Dr. Perlman provided valuable feedback on how the socioeconomic factors could potentially influence my analyses and also on how to strengthen the discussion of the results. I thank both of you for your valuable time and guidance. It was an absolute pleasure learning from you.

Third, I would like to thank Dr. Ashraf Dewan, my former supervisor at the undergraduate level, for introducing me to the world of quantitative and spatial research. Without your valuable support, I would not have achieved the goal of studying at the University of Waterloo.

I would also like to thank Prof. Dr. Towhida Rashid, for supporting me throughout my Bachelor's degree at the University of Dhaka and for believing in me during my difficult times. I want to express my sincerest gratitude to Mr. Atique Iqbal Chowdhury, for being an exceptionally co-operative boss and giving me the freedom while working at icddr,b, which ultimately helped me grow as an independent researcher. I would also like to thank Mr. S.K Masum Billah for inspiring me to pursue a higher degree in public health.

I especially want to thank Mrs. Shama Khan and Mr. Imtiaz Karim Naim, for being nothing less than guardians for me in Canada. I truly cannot express enough gratitude to you for always making me feel that I have a family here, away from home. I am extremely grateful to Mr. Durjay Chowdhury and Mr. Tauhid Khan for being two exceptional elder brothers. You two have given me some really cherishable memories. Also, thank you, Mr. Nazmus Sakib, for extending your help and support when I was lost after landing in Canada. I would also like to thank Ms. Katelyn Folkerts for introducing me to various Canadian cultures and for inviting me on your family occasions.

I express my heartfelt gratitude to all my friends at the University of Waterloo and the staffs and faculty members at the School of Public Health and Health Systems.

Finally, and foremost, I want to thank my parents, who have sacrificed all the luxuries and comforts to bring me where I am today. I would have never understood the importance of you two in my life if I had not lived alone in Canada for two long years. I am lucky to have received your unconditional love and affection. Also, cheers to my brother, Yousha and my sister Maria, for keeping my life vibrant, and for being the shoulders to lean on when I am at my worst.

Table of Contents:

List of Figures:

List of Tables:

List of Abbreviations:

LULC	Land Use/Land Cover
NDVI	Normalized difference vegetation index
EVI	Enhanced vegetation index
SAVI	Soil-adjusted vegetation index
BSM	Bayesian Spatial Modeling
ICES	Institute for Clinical Evaluative Sciences
MOHLTC	Ontario Ministry of Health and Long-Term Care
CAPE	Client Agency Provider Enrollment
OHIP	Ontario Health Insurance Plan
OLI-TIRS	Operational Land Imager and Thermal Infrared Sensor (Landsat)
TOA	Top-of-atmosphere
RF	Random Forest
LiDAR	Light Detection and Ranging
UTM	Universal Transverse Mercator
OMI	Ontario Marginalization Index
ICAR	Intrinsic Conditional Autoregressive
DIC	Deviance Information Criterion
ADHD	Attention Deficit Hyperactivity Disorder

Chapter 1 Background

1.1 Vegetation and mental health

The influence of vegetation on mental health is a topic of considerable debate in recent times. Several carefully designed studies have obtained contradictory results while assessing the role of green space and vegetation in improving mental health conditions. Evidence suggests that the vegetation-covered areas can help improve mental health, particularly through physiological stress relief (Annerstedt et al., 2013; Brown, Barton, & Gladwell, 2013), building attention restoration capacity (Mitchell & Popham, 2008), and promoting social cohesion (Markevych et al., 2017). Contrary to these findings, some studies also reported that vegetation-covered green spaces are either not associated or weakly associated with mental health (Ma, Li, Kwan, & Chai, 2018; Melis, Gelormino, Marra, Ferracin, & Costa, 2015). These studies argue that the influence of vegetation is inconsistent across different study groups to conclude any significant association with mental health. Additionally, their findings suggest that people living in densely vegetation-covered areas could already be socioeconomically advantaged, which instead of vegetation could be the actual determinant of mental well-being.

Due to the existing disagreements amongst researchers on the importance of vegetation in mental health, the effect of vegetation on mental health disorders has remained mostly unexplored. Few studies that studied this relationship have found that vegetation can positively affect patients with severe mental health disorders, such as affective and psychotic disorders (Bielinis, Jaroszewska, Łukowski, & Takayama, 2020; Chen, Yu, & Lee, 2018). When patients with affective disorders were treated with forest therapies, where the patients had engaged in recreational activities in the nearest suburban forest (with dense vegetation cover), positive effects on 'confusion' and 'depression' were noticed. Similarly, for patients with psychotic disorders, there were significant improvements in the major symptoms of schizophrenia, such as anxiety, dejection and confusion. Interestingly, research has also shown that simple forest bathing in the presence of a large number of trees might not yield the desired psychological health benefits (Takayama, Saito, Fujiwara, & Horiuchi, 2017). Instead, a properly managed vegetation-covered area, where people can tangibly and consistently experience the surrounding greenness, would have a notable impact on people's psychology (Markevych et al., 2017; Takayama et al., 2017).

Although age and sex differences are evident in mental health disorder cases (Bangasser & Valentino, 2014; Feehan, McGee, & Williams, 1993; Jones, 2013; Morgan, Castle, & Jablensky, 2008), whether these differences impact the association with vegetation have not been studied thoroughly. Dadvand et al. (2016) conducted a survey on 3461 adults in Barcelona and found that the surrounding residential vegetation or greenness was positively associated with better mental health status for only males younger than 65 years (Dadvand et al., 2016). However, Astell-Burt et al. (2014) reported that vegetation-covered green space was associated with better mental health for both males and females. Their study found that the benefit of green space on mental health was evident for young to middle-aged (30-45 years) men, while only older (≥ 45 years) women were benefitted from exposure to greenness (Astell-Burt, Mitchell, & Hartig, 2014). In contrast, Villeneuve et al. (2018) could not find any age and sex-specific differences in the association between surrounding greenness and mental health conditions of the study participants (Villeneuve et al., 2018). Therefore, the results from these studies considerably differ from one another and indicate a need for further research. It is imperative to address the possible model constraints and the limitations in existing modeling techniques to elucidate the differential influence of vegetation on the mental health conditions of males and females in different age groups.

Despite the contradictory findings of past studies, the relationship of vegetation with the mental health of the general public remains an issue of considerable interest. This is primarily because the global increase in urbanization has given rise to some unique environmental problems, such as the substantial loss of vegetation covered areas, which might adversely affect the mental health of people, particularly the urban dwellers. For example, Martellozzo et al. (2015) reported that owing to the urban and peri-urban growth in the Calgary-Edmonton corridor in Canada, a vast expanse of vegetation was cleared away (Martellozzo et al., 2015). Similarly, during the early stages of urbanization at southeastern Wisconsin in the USA, a considerable loss of natural and semi-natural vegetation was accompanied by the growth of the city (Sharpe, Stearns, Leitner, & Dorney, 1986). Around 80% of the total population in North America currently live in urban areas and by 2050, around 70% of the world population will live in similar urban settings (United Nations, 2018). Therefore, the role of vegetation in improving the mental health conditions of the mass people must be understood immediately, and necessary planning strategies should be devised to ensure the mental well-being of the urban population.

1.2 Challenges in selecting the appropriate vegetation measure

There are several challenges in studying the relationship between vegetation and mental health. One of the main challenges stems from the fact that characterizing "vegetation" in mental health studies can be extremely difficult since vegetation comes in different forms, including tall trees in protected areas, shrubs and bushes in parks, and ornamental plants in gardens and roof-tops. Therefore, the association between vegetation and mental health could show differential sensitivity depending on the type of vegetation measure (Markevych et al., 2017). Ideally, the best vegetation measure will be the one that can effectively capture all forms of vegetation and people's perception of surrounding greenness in the study area (Dzhambov et al., 2018b; Jiang et al., 2017; Markevych et al., 2017).

Rugel et al. (2017) discussed that remote sensing-based indices, such as the normalized difference vegetation index (NDVI), could be used to effectively characterize vegetation in population-level mental health research (Rugel, Henderson, Carpiano, & Brauer, 2017). The findings of past studies, where higher NDVI values were found to be significantly associated with better mental health conditions, support this claim (Dzhambov, 2018; Dzhambov et al., 2018b). Interestingly, some studies have employed NDVI and have obtained a non-significant association between vegetation and mental health. For example, Villeneuve et al. (2018) studied the association between NDVI and mental health conditions of 282 adults in Ottawa, Canada, and found that the NDVI was not associated with the mental health of the study participants (Villeneuve et al., 2018). Similarly, a large study in South Africa, comprising 11150 participants, had analyzed the relationship between the green environment and the incidence of depression and found that the health benefits of NDVI are uneven and strata-specific (Tomita et al., 2017). The study found that the association between NDVI and depression was non-significant at the population level but was significant for the depression of the middle- and low-income participants.

Similar to NDVI, discordant findings are also evident for composite or area-based measures of vegetation. For example, Huynh et al. (2013) extracted the vegetation and water body covered areas, using circular buffers, and classified them as "natural space" to study their influence on the emotional well-being of young people in Canada (Huynh, Craig, Janssen, & Pickett, 2013). Their study found that the relationship between natural space and the positive well-being of young people were weak and inconsistent. On the contrary, Rugel et al. (2019) used a similar buffer- and polygon-based measure of natural space and discussed that the accessibility to natural spaces could yield better mental health outcomes for people in dense urban areas, when mediated by a higher sense of community belonging (Rugel, Carpiano, Henderson, & Brauer, 2019).

In this regard, Markevych et al. (2017) discussed that in population-based health studies, highly sensitive vegetation indices such as enhanced vegetation index (EVI) (USGS, 2019a) and soil-adjusted vegetation index (SAVI) (Huete, 1988) could be more useful, compared to the conventionally used NDVI and polygon-based measures of vegetation (Markevych et al., 2017). Unlike NDVI, these indices can better capture vegetation signals in various built-up and climatic settings by adjusting for the atmospheric disturbances, background canopy cover, and spurious soil brightness (Jensen, 2009; USGS, 2019a, 2019b). Although these vegetation indices cannot differentiate between structured (vegetation at formal settings like parks) and unstructured (vegetation around homes, backyards and street-sides) forms of vegetation, these indices can give a good relative measure of the health or quality of vegetation patches. Therefore, in contrast to polygon and area-based vegetation measures, where only the areal extent of the vegetation is available, these indices can directly estimate the relative exposure to vegetation or the greenness that an individual perceives.

Most importantly, Jiang et al. (2017) noted that vegetation measures that can best capture people's perception of surrounding greenness should always be used for policymaking purposes (Jiang et al., 2017). This issue is often overlooked by present-day epidemiological studies, which aim to study the effect of vegetation on mental health to guide policymaking. Additionally, Markevych et al. (2017) discussed that perceived greenness could be a better measure for studying the association of vegetation with health outcomes. This is because the perceived greenness directly corresponds to the quantity and quality of people’s exposure to the surrounding greenness. Markevych et al. (2017) had also strongly emphasized the need for future research to analyze whether the association with health indicators, such as mental health, show differential sensitivity to the various satellite-derived indices. Thus, these studies highlighted a vast research gap, which could be addressed through a study analyzing the association between mental health and different indices- and area-based measures of vegetation.

1.3 Challenges in selecting the appropriate modeling technique

The distributions of mental health disorder cases and vegetation cover are both spatially structured due to their varying levels of distribution in space (Gould, 2000; Sheppard et al., 2012). Therefore, a careful selection of a statistical model is necessary to analyze the association between these two variables. The traditional and non-spatial statistical models, such as multiple linear or logistic regression models assume structural stationarity of variables over space, which can be a gross oversimplification of the real-life scenario (Anselin, 1995). In particular, when cases of mental health disorders or vegetation cover systematically vary across space or are spatially dependent on the neighboring values, this oversimplification may lead to the violation of core model assumptions that the observations and the residual errors are independent from each other or do not show any interdependence (Anselin, 1990; LeSage, 1997). Consequently, spatial dependence may impair the estimation of beta (β) coefficients and the accuracy of a significance test, affecting both the magnitude and significance of the association between vegetation and mental health.

It is also important to note that not all spatial models can capture the underlying (latent) data generating processes (Robertson, Nelson, MacNab, & Lawson, 2010). These processes can be a group of socioeconomic and sociocultural factors, which the researchers are unable to measure and incorporate in the models but have played a vital role in determining the distribution of the dependent and independent variables (Law & Perlman, 2018). Without incorporating these latent processes in regression models, the true effect of vegetation on mental health cannot be measured precisely. Some frequentist spatial models, such as the spatial error and spatial lag models, consider spatial effects as a nuisance and adjust them accordingly during the estimation of β coefficients (LeSage, 1997; Robertson et al., 2010). As a result, using these frequentist models, it is not possible to measure the relative contributions of the unmeasured spatial and non-spatial covariates in the data generating process (Law & Haining, 2004; Law, Haining, Maheswaran, & Pearson, 2006; Law & Perlman, 2018). However, to understand how humans interact with vegetation in space (spatial process) and how these interactions, in turn, influence the mental well-being of the study population (non-spatial process), it is essential that epidemiological studies attempt to understand the dynamics of the latent spatial and non-spatial processes (Dzhambov, 2018; Dzhambov et al., 2018b; Rugel et al., 2017).

Contrary to the frequentist approaches, Bayesian Spatial Modeling (BSM) can be applied to capture both the spatial dependence and spatial structure of the covariates through the integration of a spatial random effect term (s_i) in the models. Additionally, any overdispersion in the count data of mental health disorders could be adjusted using a non-spatial or spatially unstructured random effect term (u_i) (Law & Haining, 2004; Law et al., 2006). Furthermore, epidemiological studies are often interested in the area wise relative risk of mental health disorders, which cannot be estimated precisely when the population size is too small or large. For example, extreme relative risk values are commonly associated with areas having small populations, while statistically significant relative risk values are associated with areas with large populations (Law et al., 2006). These artifacts owing to variations in the population size can also be adjusted in Bayesian models through the process of 'borrowing' information from adjacent areas. Under this process, the models carry out statistical smoothing and incorporate the prior information (evidence from the data of surrounding areas) and the observed data of the area for which the risk will be estimated. Therefore, any statistical artifacts such as small data counts and large variations in sample size or study populations are inherently adjusted in the Bayesian models (Law & Haining, 2004; Law et al., 2006; Lawson, 2013). Consequently, through the application of BSM, it is possible to accurately identify areas with high risk of mental health disorders due to the influence of putative risk factors like low vegetation content.

Chapter 2 Study aims and objectives

2.1 Study rationale and contribution to the knowledge gaps

Existing works from the available literature suggest that there is a need to understand whether the association of mental health disorders with vegetation can show differential susceptibility to various measures of vegetation. Additionally, past studies suggest that it is essential to investigate whether the beneficial effects of vegetation cover on the mental and psychological well-being of the urban population can vary according to the age and sex structures of the urban population. This study attempted to address these two distinct research gaps using geostatistical techniques.

The most notable contribution of this study in the public health domain is to establish a comparison of the effect of integrating different measures of vegetation to study the association between vegetation and mental health disorders. The study is expected to bridge the knowledge gap by answering which type of vegetation measure could be potentially suitable for population-based epidemiological studies, which aim to assess the exposure of the general public to the surrounding vegetation.

2.2 Aims and objectives

The primary objective of the thesis is to understand the differential impact of vegetation measures on the association between vegetation and mental health disorders.

The three specific aims of the thesis are:

1. Understanding how indices- and area-based measures of vegetation can impact the association between vegetation and different forms of mental health disorders, specifically, the psychotic and non-psychotic disorders.

2. Identification of a suitable vegetation measure to study the association between vegetation and mental health disorders.

3. Analyzing the association between vegetation and the age and sex-specific mental health disorders.

Chapter 3 Methods

3.1 Study area

This study focused on the City of Toronto, with a population of 2,731,571 in 2016 (Canadian census, 2016). The study was conducted at the neighborhood level and all the 140 neighborhoods were considered for the analysis. The neighborhoods were defined by the Social Policy Analysis and Research Unit in the Social Development and Administration Division of the City of Toronto (Law & Perlman, 2018). These neighborhoods are geographic units created for planning and service delivery purposes by aggregating the Statistics Canada Census Tracts into meaningful spatial units (Toronto Community Health Profiles).

The City of Toronto is one of the most urbanized and populous cities of the Ontario province in Canada and due to the high urbanization rate, the built environment is becoming the dominant land cover type in the area. The proliferation of the built environment is believed to have drastically reduced the soil volume available for the growth of small vegetation and has also decreased the aerial space for the growth and expansion of large trees. The increase of built-surfaces has led to a rise in non-permeable surfaces and the ground salinity level due to the use of de-icing salt in the roads during winter seasons. The environmental impacts of these changes are the dehydration and death of natural flora (City of Toronto, 2013).

3.2 Data preparation

3.2.1 Mental health disorders

The mental health disorder data, covering the period from 1st April 2015 to 31st March 2016 (Fiscal year, 2015), were retrieved from the Ontario Community Health Profiles Partnership database (Ontario Community Health Profiles Partnership, 2019). The dataset was a part of the study, "Enrollment, Access, Continuity and Mental Health Gaps in Care (ICES Project No. 2018 09000 992 000)", which was supported by the Institute for Clinical Evaluative Sciences (ICES) and funded by the Ontario Ministry of Health and Long-Term Care (MOHLTC). Further details on the project and the inclusion and exclusion criteria could be found in *Appendix A*.

For the first and the second aims of the study, the data on the psychotic and non-psychotic disorder cases were utilized. The observed counts on psychotic and non-psychotic disorder cases for both sexes (males and females) and ages 0+ years were used to understand how indices- and area-based measures of vegetation can impact the association between vegetation and the psychotic and non-psychotic disorders. The original data were divided into 'enrolled' and 'non-enrolled' categories, where the term 'enrolled' relates to the primary care enrollment models that are found in the Client Agency Provider Enrollment (CAPE) tables. The CAPE tables are used to identify patients enrolled in different primary care models over time (Glazier et al., 2018). As the target was to model the distribution of the mental health disorder cases, regardless of the patients' enrollment statuses in the primary care models, both enrolled and non-enrolled cases were combined for the analyses. The final dataset contained counts of all the Ontario permanent residents with an Ontario Health Insurance Plan (OHIP) and having OHIP claims for the mental health conditions listed in Table 1 (Glazier et al., 2018).

The results of these analyses were used to select one single vegetation measure for analyzing the association between vegetation and the age and sex-specific mental health disorders. The exact selection criteria are detailed in Section 3.2.7.

Table 1: Categories and sub-categories of mental health disorders and OHIP codes

Type	**Sub-category**	**OHIP codes of sub-category**
1) Psychotic disorders	Schizophrenia	295
	Manic-depressive psychoses, involutional melancholia	296
	Other paranoid states	297
	Other psychoses	298
2) Non-psychotic disorders	Anxiety neurosis, hysteria, neurasthenia, obsessive-compulsive neurosis, reactive	300
	Personality disorders	301
	Sexual deviations	302
	Psychosomatic illness	306
	Adjustment reaction	309
	Depressive disorder	311
3) Substance-use disorders	Alcoholism	303
	Drug dependence	304
4) Family, social and occupational issues	Economic problems	897
	Marital difficulties	898
	Parent-child problems	899
	Problems with aged parents or in-laws	900
	Family disruption/divorce	901
	Education problems	902
	Social maladjustment	904
	Occupational problems	905
	Legal problems	906
	Other problems of social adjustment	909
Combined mental health disorders	Psychotic, non-psychotic, substance-use and family, social and occupational issues related disorders	(All codes listed above)

The observed count data for the combined mental health disorder variable (Table 1) were used to complete the third aim of the study, more specifically, to analyze the association between vegetation and the age and sex-specific mental health disorder cases. The data for combined mental health disorders was age and sex-stratified and contained four types of disorders, namely, psychotic, non-psychotic, substance-use, and disorders related to family, social and occupational issues. More details on the four major types of disorders could be found in Table 1. The combined mental health disorder data were grouped into four age-groups, 0-19, 20-44, 45-64 and 65+ years.

The expected counts of the mental health disorders were derived separately for the psychotic, non-psychotic, and the combined mental health disorder cases using an indirect (internal) standardization method. The expected counts of psychotic, non-psychotic, or combined mental health disorders correspond to the overall rate of these disorder cases multiplied with the residential population of each of the neighborhoods. This process involved applying the age and sex-specific rates to the population structure of each neighborhood and calculating the expected number of cases. As the age of the individuals was not used to group the data for psychotic and non-psychotic disorder cases by the data provider, only the sex-specific rates could be used to estimate the expected counts of psychotic and non-psychotic disorder cases.

The quantile maps in Figure 1 and Figure 2 show that both the high and low rates of mental health disorders are concentrated at particular parts of the Toronto area, suggesting that the mental health disorder cases could be spatially autocorrelated. Since it is highly necessary to confirm the presence or absence of spatial autocorrelation in the data prior to the selection of an appropriate modeling technique, the global Moran's I test was carried out using the GeoDa software (https://geodacenter.github.io/). As hypothesized, the results suggested the presence of significant spatial autocorrelation in the data, which indicated that a spatial modeling technique would be essential to study the association between vegetation and the different types of mental health disorder cases. The details of the spatial autocorrelation test could be found in *Appendix B.*

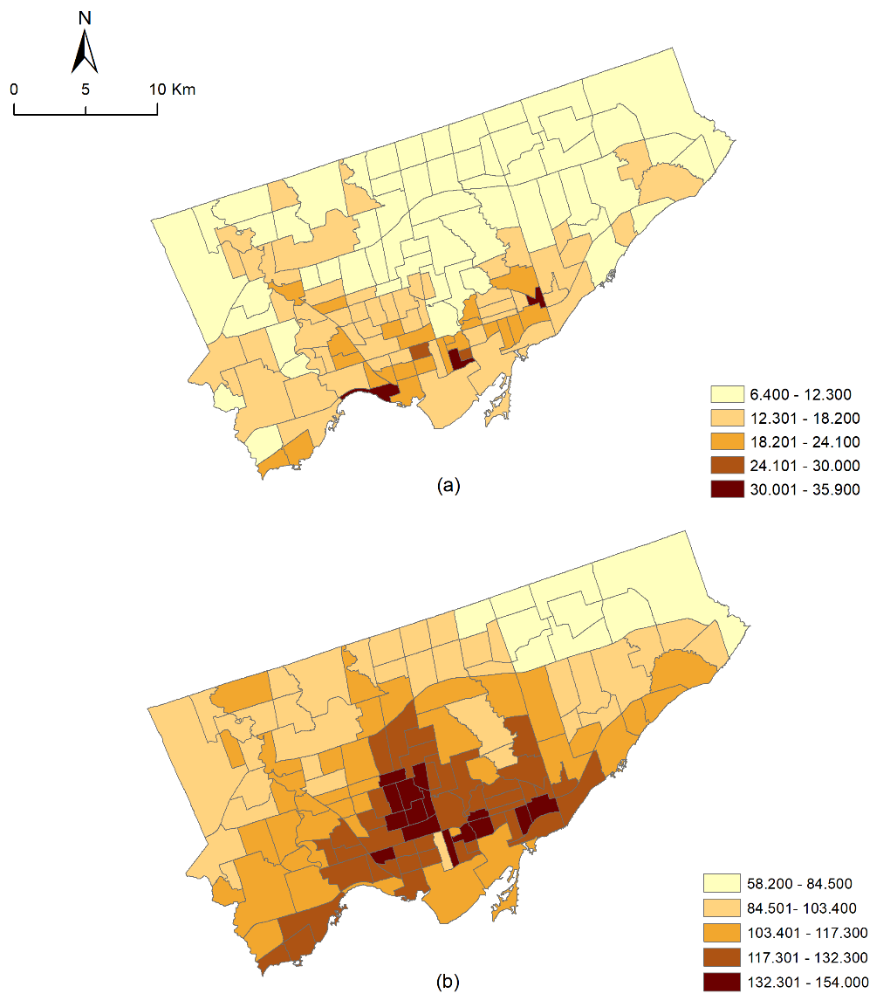


Figure 1: Quantile maps (at equal intervals) showing the age and sex-standardized rates of a) psychotic and b) non-psychotic disorders in the City of Toronto

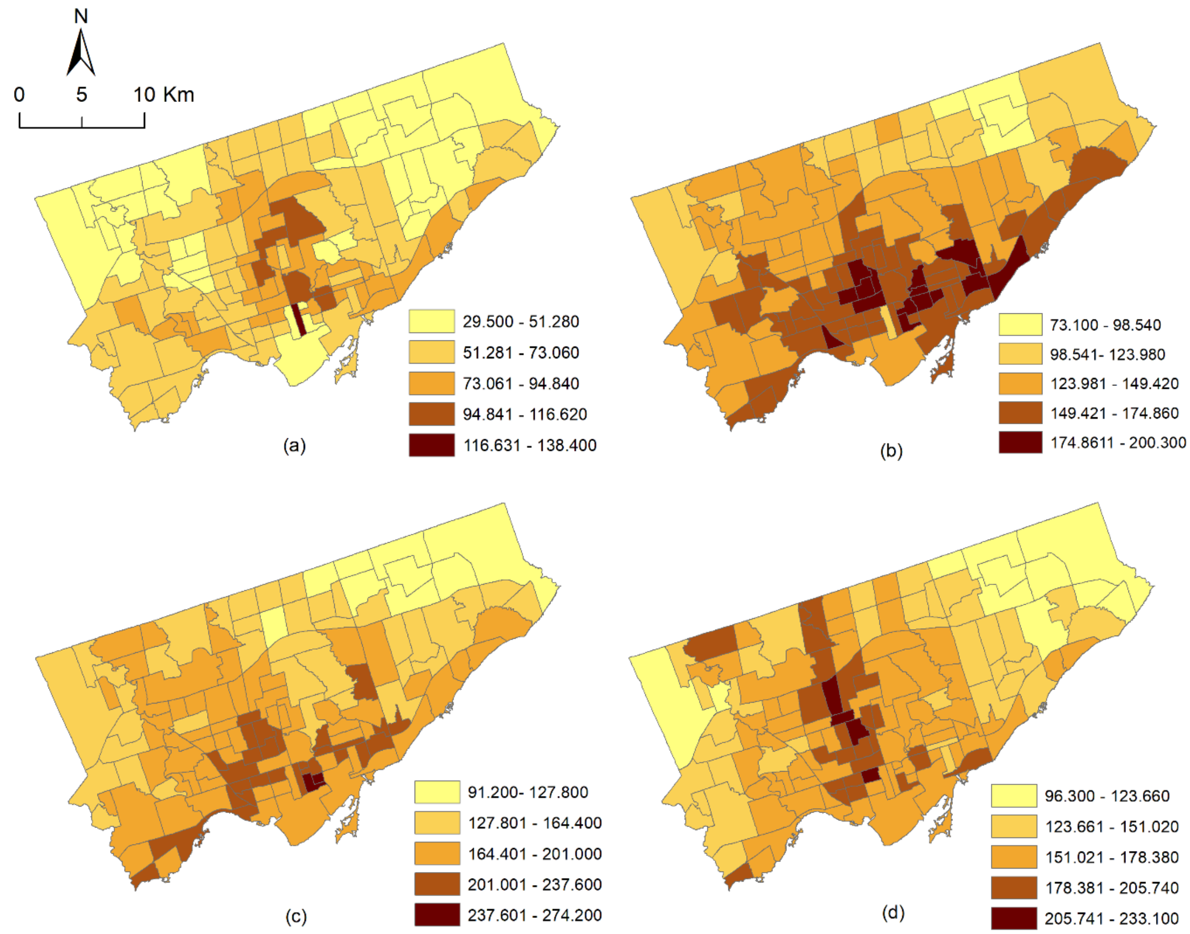


Figure 2: Quantile maps (at equal intervals) showing the crude rates of the combined mental health disorders for both sexes and the age groups (a) 0-19, (b) 20-44, (c) 45-64 and (d) 65+ in the City of Toronto

3.2.2 Landsat 8 satellite imageries

Three satellite images were retrieved from the Landsat Operational Land Imager and Thermal Infrared Sensor (OLI-TIRS) or Landsat 8 from the USGS EarthExplorer data repository (USGS-EarthExplorer). A search criterion of less than 10% cloud cover was used to exclude the images with a considerable presence of clouds since cloud cover in satellite images can considerably influence the calculation of the vegetation indices (Jensen, 2009). Three images, having an average cloud cover of 2.67% and a spatial resolution of 30 m, were required to cover the study area. Two of these images were acquired on 20th May 2016 and the third image was captured on 14th June 2016. The year 2016 was selected to be consistent with the data period of mental health disorders. Similarly, the May-June months were chosen to estimate the vegetation content of the spring-summer seasons because the vegetation densities during these months become stable after a cold-snowy winter and before a chilly fall season.

Radiometric corrections were conducted by converting the raw digital numbers to the top-of-atmosphere (TOA) reflectance (USGS, 2017). Furthermore, atmospheric corrections were applied to remove any haze from the images through the identification of the darkest pixel value in each band and subtracting this value from every pixel in the satellite image (Song, Woodcock, Seto, Lenney, & Macomber, 2001). Finally, the radiometrically and atmospherically corrected images were mosaiced and cropped using the boundary of the Toronto city to produce a single image for the analysis.

3.2.3 Construction of the vegetation indices

The three vegetation indices, EVI, NDVI, and SAVI, were generated using the processed Landsat 8 image for Toronto. The details of the vegetation indices used in this study and the computational formulas that were used are tabulated in Table 2. The *Raster Calculator* in ArcMap 10.7 software (https://desktop.arcgis.com/en/arcmap/) was used to perform the band operations to produce the rasters of vegetation indices. The formulas for the computation of these vegetation indices were obtained from the USGS websites for each of the indices (USGS, 2019a, 2019b, 2019c).

These three indices are computed from different ranges of wavelengths in the electromagnetic spectrum (referred to as bands), reflected from the vegetation surface and received by the satellite. The reflectance of these bands, in turn, is governed by factors such as the type of plant, water retention capacities of the tissues, chemical and morphological characteristics of the leaves, and the level of photosynthetic activities in the plant (Peñuelas & Filella, 1998; Xue & Su, 2017; Zhang & Kovacs, 2012). Consequently, a number of remote sensing-based vegetation indices have been proposed, each derived from a different set of spectral bands and suited to capture the vegetation content in an area based on the climatic, physical and geomorphic characteristics (Jensen, 2009; Xue & Su, 2017).

Finally, the NDVI raster was used to extract the vegetation-covered areas and to mask out the non-vegetation features like water body, bare soil and built-up surfaces in all of the three vegetation rasters. This process was necessary to remove the negative values representing the non-vegetation features in the vegetation indices (Markevych et al., 2017). Finally, the mean values of the indices for each neighborhood were extracted using the *Zonal Statistics* tool in ArcMap 10.7.

Table 2: Details of the vegetation indices used in this study

Vegetation Indices	Formula	Description
Enhanced Vegetation Index (EVI)	**Generic:** EVI = G * ((NIR - R) / (NIR + C1 * R – C2 * B + L_{EVI})) **For Landsat 8:** EVI = 2.5 * ((Band 5 – Band 4) / (Band 5 + 6 * Band 4 – 7.5 * Band 2 + 1))	EVI is a vegetation index that quantifies the vegetation greenness. Compared to other similar indices, EVI adjusts atmospheric conditions and canopy background noise and is more sensitive in areas with dense vegetation (USGS, 2019a) The higher the value of EVI, the greater is the vegetation content and the greenness of the area.
Normalized Difference Vegetation Index (NDVI)	**Generic:** NDVI= (NIR - R) / (NIR + R) **For Landsat 8:** NDVI = (Band 5 – Band 4) / (Band 5 + Band 4)	NDVI is the most commonly used vegetation index in health research and can help estimate the greenness or the quality of vegetation cover (USGS, 2019b). However, in contrast to EVI, NDVI cannot adjust for the atmospheric conditions and canopy background noise.
Soil Adjusted Vegetation Index (SAVI)	**Generic:** SAVI= ((NIR - R) / (NIR + R + L_{SAVI})) * (1 + L_{SAVI}) **For Landsat 8:** SAVI = ((Band 5 – Band 4) / (Band 5 + Band 4 + 0.5)) * (1.5)	Although SAVI is similar to NDVI, it can adjust for the influence of the soil brightness, which otherwise affects the estimation of NDVI in areas where the vegetation cover is low (USGS, 2019c).
Band descriptions		
NIR is the Near Infrared band of the satellite image R is the Red band of the satellite image B is the Blue band of the satellite image G is the gain factor that makes EVI comparable to that of other vegetation indices such as NDVI L_{EVI} is a constant used to adjust for the canopy background L_{SAVI} is the soil brightness correction factor C1 and C2 are constants used to adjust for atmospheric resistance		

3.2.4 Developing the land use/land cover (LULC) model using the Random Forest ensemble

A land use/land cover model was developed to estimate the percentage of vegetation-cover in the Toronto area. This LULC model was developed to compare the area-based measures of vegetation derived through the application of an advanced machine learning ensemble with the area-based measures derived using automated extraction of features through custom made procedures (such as the tree cover dataset from the Toronto Open Data Portal). Furthermore, this LULC model would also allow a comparison between the vegetation indices that are able to estimate the plant biomass vigor and the area-based measures that are simply able to measure the areal extent of vegetation cover.

The Random Forest (RF) classifier was chosen to develop the LULC model because RF is one of the most powerful machine learning classifiers to date. The RF classifier can handle the classification of both multispectral and hyperspectral satellite images in noisy, unbalanced and non-linear data settings (Abdullah et al., 2019; Breiman, 2001; Cutler, 2004). The use of the RF ensemble ensures significantly better classification accuracies, especially when the classification for areas such as Toronto could be heavily complicated due to the mixture of built-environment and natural features (Gislason, Benediktsson, & Sveinsson, 2006; Puissant, Rougier, & Stumpf, 2014). For example, it would be extremely challenging for an algorithm to distinguish between a green-colored building and a tree with a large green canopy.

This study employed the 'randomForest' package in R for the classification process (https://cran.r-project.org/web/packages/randomForest/randomForest.pdf). Google Earth aerial imagery for the Toronto area in 2016 was used to generate the training data to be later used in the RF algorithm. On-screen visual interpretation and the NDVI image were used to assist the generation of training dataset for the vegetation class. A total of 400 training data points were generated to develop the LULC model. A uniform number of training data (100 per class) was maintained for all the four land cover classes listed in Table 3. Furthermore, to ensure a homogenous distribution of training dataset over the study area, all the parts (north, south, east and west) of the study area were equally considered for generating the training data. Finally, to assess the accuracy of the developed model, 25% of the training data were retained for accuracy assessments, which means 75% of the data was used for training the RF model.

Table 3: Land cover classes developed in this study

LULC Types	Description
Bare soil	Exposed soils, construction sites
Built-up	Residential, commercial and services, industrial, transportation, roads, mixed urban, and other urban
Vegetation	Deciduous forest, mixed forest lands, palms, conifer, scrub, and others
Waterbody	Permanent and seasonal wetlands, inland water bodies, low-lying areas, marshy land, rills and gully, swamps

A different number of input features (*mtry*) and the number of decision trees (*ntree*) parameterization were performed to inspect the out-of-bag (OOB) error rates. Finally, an RF model was trained with a *ntree* and *mtry* setting that contributes to the lowest OOB error rate. Consequently, the final RF model on the training data was trained using the number of decision trees (*ntree*) as 500 and the number of input features (*mtry*) as 3. This trained model was then used for predicting LULC classes in the satellite image. Lastly, the vegetation-covered areas were extracted from the LULC raster and the '*Tabulate Intersection*' tool in the ArcMap software was used to estimate the percentage of area covered by vegetation (hereinafter referred to as Veg_RF) in each neighborhood.

3.2.5 Processing the tree cover dataset from Open Data Portal

The tree cover dataset was retrieved from the 'Treed area' data in the Toronto Open Data Portal (Open Data Portal Toronto, 2019). The City of Toronto's Open Data Portal is a public data repository, which allows developers, students and researchers to easily avail spatial and non-spatial datasets related to the functioning of the city. The tree cover dataset was developed via automated extraction from aerial Light Detection and Ranging (LiDAR) using custom-developed procedures and open source tools and was a representation of the physical features (trees) that were visually identifiable in an aerial photograph (Open Data Portal Toronto, 2019). The data was downloaded in a shapefile (.shp) format and was converted to the Universal Transverse Mercator (UTM) projection for further use. The percentage of area covered by trees in each neighborhood (hereinafter referred to as Tree_OD) was estimated using the *Tabulate Intersection* tool in ArcMap.

3.2.6 Adjusting for potential confounders

The socioeconomic factors can have a profound impact on the mental well-being of people of all ages and sexes (Mckenzie, Gunasekara, Richardson, & Carter, 2014; Reiss, 2013; Saraceno, Levav, & Kohn, 2005). For example, psychotic disorders like schizophrenia were found to be more prevalent in the lower than in higher socioeconomic groups (Saraceno et al., 2005). Several socioeconomic factors such as social discrimination, unemployment and poverty-related stress are believed to be influential factors for the occurrence of these disorders.

More specifically, factors such as material deprivation, residential instability, dependency and ethnicity have well-documented influences on mental health conditions (Bjarnason & Sigurdardottir, 2003; Kataoka, Zhang, & Wells, 2002; Sasaki, Vega, & McGowan, 2013; Satcher, 2001). For example, a significant decline in the self-reported mental health conditions was found in New Zealand due to the increases in individual deprivation (Mckenzie et al., 2014). Similarly, a longitudinal study conducted on 2754 Canadians using data from the Canadian National Population Health Survey, had found that worsening material deprivation is associated with the self-reported psychological distress of study participants (Blair, Gariépy, & Schmitz, 2015). Material deprivation is directly related to poverty and, therefore, represents the economic constraints that prevent people from attaining basic materials for sustenance. Due to this, psychological stress accumulates over time, which eventually takes into the form of mental health disorders (Mckenzie et al., 2014; Saraceno et al., 2005).

The ethnic concentration in Canada is also an important factor to be considered for mental health studies because past research on immigration in Ontario had shown that the new immigrants in Canada have a 'healthy immigrant effect' and help improve the overall health conditions in the region (Khan et al., 2017; Matheson FI & van Ingen T, 2018). However, a large cross-sectional study conducted on 10,000 non-institutionalized residents in Spain reported that when employment and material deprivation were kept unchecked, the overall health conditions of immigrants could be severely impacted (Borrell et al., 2008). Additionally, racism and discrimination towards ethnic groups were found to adversely affect the mental health conditions of people (Pieterse, Todd, Neville, & Carter, 2012). Therefore, it is necessary to incorporate ethnic concentration to study mental health as poor social and economic conditions of ethnic people could constitute poor mental health conditions in the study area.

The socioeconomic covariates were retrieved from the Ontario Marginalization Index (OMI) (Matheson FI & van Ingen T, 2018) to be adjusted as potential confounders in the models. The OMI comprises of four major dimensions or categories, these are:

1) **Material Deprivation:** This dimension was created from the indicators that measure income, quality of housing, education attainment and family structure characteristics such as family who are lone-parent families. As explained earlier, material deprivation is directly related to poverty and people's capacity to access and avail basic necessities.
2) **Residential Instability:** This dimension was constructed from the indicators that measure the types and density of residential accommodations and certain family structure characteristics such as the proportion of the population who are single, divorced or widowed. Residential instability captures the quality of neighborhoods, cohesiveness and supports in terms of these indicators.
3) **Dependency:** This dimension originated from the indicators that measure the area-level concentrations of people who are not compensated for their work or who do not receive income from employment. This group comprises of seniors, children and people with disabilities.
4) **Ethnic concentration:** This dimension was made from the indicators that measure high area-level concentrations of people who are recent immigrants and people who belong to a visible minority group

A detailed list of the indicators used to create each of the four variables could be found in *Appendix C*.

In order to avoid multicollinearity due to the addition of these four socioeconomic variables, Pearson correlation coefficient (Benesty, Chen, Huang, & Cohen, 2009) and multicollinearity (Mansfield & Helms, 1982) tests were conducted to check whether they were significantly correlated and whether these dimensions could be linearly predicted from one another. The results of the tests indicate that the OMI variables do not demonstrate sufficient intercorrelations and multicollinearity and, therefore, all four of the variables could be included in a regression model. The details of the correlation and multicollinearity tests are presented in *Appendix D.*

The weighted average scores for the variables were used in this study, where a high score represents high material deprivation, residential instability, dependence or ethnic concentration. In this regard, high ethnic concentration implies a high concentration of recent immigrants and visible minorities (Matheson FI & van Ingen T, 2018).

In addition to the socioeconomic factors, substance use disorder may have a marked effect on the mental well-being of people, aggravating mental conditions such as anxiety, depression and even dementia (Han, Gfroerer, Colliver, & Penne, 2009; Reid & Anderson, 1997). Older adults were found to be more affected by substance use disorders and the associated mental health complications than young people (Simoni-Wastila & Yang, 2006). Currie et al. (2005) analyzed data from the Canadian Community Health Survey and found that the substance use disorder co-occurs in high frequency in cases of major depressive disorders (Currie et al., 2005). Their study also found that substance dependence can help predict the higher prevalence of suicidal thoughts and mental health treatment use in adults. Therefore, to adjust for the effect of substance use disorder on mental health disorders, the age and sex standardized rate of substance use disorders (both sexes, 0+ age and per 1000 population) was retrieved from the Ontario Community Health Profiles Partnership database (Glazier et al., 2018; Ontario Community Health Profiles Partnership, 2019) and added as a potential confounder in the Bayesian models.

The substance use disorder variable was only added as a confounder to study the association of different vegetation measures with psychotic and non-psychotic disorders (Study

Aim 1). This variable was not added as a confounder to study the age and sex-specific effect of vegetation on mental health disorders (Study Aim 3) since the outcome variable, the combined mental disorder cases, already contained the substance use disorder data (Table 1).

The family, social and occupational issues variable was not adjusted in any of the models in this study because the effect of the family, social and occupational issues were adequately captured by the OMI variables. Table 1 shows the sub-categories for the family, social and occupational issues and it could be observed that the sub-categories are very similar to the indicators (*Appendix C*) used to construct the four OMI dimensions.

The summary statistics of the variables used in this study are tabulated in Table 4.

Table 4: Summary statistics of the key variables used to study the association between vegetation and mental health disorders

Variables	**Minimum**	**Mean (Standard deviation)**	**Maximum**
Dependent variables			
Number of psychotic disorders	94	282.864 (±152.637)	861
Number of non-psychotic disorders	757	2239.850 (± 964.286)	5523
Number of combined mental health disorders			
0-19 (Males)	32	111.429 (±55.039)	310
(Females)	34	114.564 (±56.387)	313
20-44 (Males)	135	417.007 (±238.416)	1669
(Females)	201	573.336 (±300.440)	2056
45-64 (Males)	126	390.557 (±182.715)	1059
(Females)	167	516.486 (±213.212)	1200
65+ (Males)	46	178.379 (±79.401)	453
(Females)	67	281.743 (±135.326)	699
Independent variables (vegetation)			
EVI	0.037	0.052 (± 0.006)	0.0679
NDVI	0.473	0.561 (± 0.035)	0.634
SAVI	0.041	0.058 (± 0.006)	0.075
Percentage of vegetation cover (Veg_RF)	0.501	20.730 (± 13.267)	54.279
Percentage of tree cover (Tree_OD)	0.100	6.540 (± 5.611)	34.117
Independent variables (others)			
Material deprivation (OMI)	-1.520	0.250 (± 0.895)	3.068
Residential instability (OMI)	-0.785	0.723 (± 0.783)	3.009
Dependency (OMI)	-1.262	-0.228 (± 0.393)	0.897
Ethnic concentration (OMI)	-0.317	0.902 (± 0.838)	3.282
Substance use disorder rate	2.410	9.988 (± 4.392)	30.54

3.2.7 Bayesian Spatial Modeling

The association between vegetation and mental health was analyzed using the Bayesian Spatial Modeling (BSM) technique. For this process, the observed counts, O_{ik}, of the mental health disorder *k*, in neighborhood *i*, was assumed to follow a Poisson distribution. In this study, *k* = 1, 2 or 3 representing psychotic, non-psychotic and the combined mental health disorder variables, respectively. Similarly, *i* = 1, 2,…n, where *n* is the total number of neighborhoods in the City of Toronto (*n* =140). Hence, Equation (1) could be used to define the distribution of the observations.

$$O_{ik} \sim Poisson(\lambda_{ik}) \quad (1)$$

where, λ_{ik} represents the expected value of the mental health disorder *k* in the neighborhood *i*

Equation (1) could be further modified to Equation (2) and (3). Equation (2) and (3) show that the observed count of mental health disorder in a neighborhood is a product of the unknown area-specific relative risk of the disorder, r_{ik}, and the expected count, E_{ik}. The E_{ik} for each neighborhood was calculated earlier using the overall rate of the disorder, *k*, multiplied with the residential population of each of the neighborhoods (as detailed in Section 3.2.1). In contrast, the r_{ik} was estimated using the Bayesian models.

Hence,

$$\lambda_{ik} = E_{ik} \text{ x } r_{ik} = E_{ik} r_{ik} \quad (2)$$

Applying logarithm to both sides of Equation (2),

$$\log[\lambda_{ik}] = \log[E_{ik}] + \log[r_{ik}] \quad (3)$$

The unknown area-specific relative risk can be assumed to be associated with the attributes of the population (socioeconomic) and environmental characteristics, or both (Law et al., 2006). As a result, for this study, the r_{ik} could be substituted by the risk owing to the area-specific variations in the vegetation content or cover.

$$\log [\lambda_{ik}] = \log [E_{ik}] + \beta_0 + \beta_1 X_{1i} \quad (4)$$

where, X_{1i} is the variable for vegetation measure (EVI, NDVI, SAVI, Veg_RF or Tree_OD)

Additionally, as noted earlier, the socioeconomic conditions (represented by the four OMI variables) and the rate of substance use can influence the observed counts of psychotic, non-psychotic and combined mental health disorders in an area. Consequently, the material deprivation (X_{2i}), ethnic concentration (X_{3i}), residential instability (X_{4i}), dependency (X_{5i}) and the age and sex standardized rate of substance use disorders (X_{6i}) were added into the model as potential confounders. Hence, Equation (4) gives,

$$\log [\lambda_{ik}] = \log [E_{ik}] + \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} \quad (5)$$

Although Equation (5) gives the desired model, there are several problems that need to be considered before finalizing the model equation. The first problem that needs to be considered is the overdispersion in count data of the observed cases of mental health disorders. One of the core assumptions of the Poisson model is that $Var\ [O_{ik}] = \lambda_{ik}$, where $Var\ [\ ]$ represents the variance. This implies that for a proper Poisson model, the mean of the observations needs to be equal to the variance of the observations. However, during overdispersion $Var\ [O_{ik}] > \lambda_{ik}$, which means the variance in the count data is higher than expected by the modeled Poisson distribution. This overdispersion stems mainly from the heterogeneity in the individual-level risk of contracting the different types of mental health disorders, which translates to the heterogeneity observed in the count data of these disorder cases. The heterogeneity in individual-level risk can be owing to the differences of individuals in lifestyles, genetic characteristics, socioeconomic and family conditions, and varying exposure to other risk factors related to poor mental health conditions. Therefore, the final model, studying the association between vegetation and mental health disorders, should always try to capture the underlying heterogeneity in the individual-level risk.

In order to adjust for the overdispersion, a Poisson lognormal model was adopted, where the individual-level processes (leading to the variations in individual-level risks) were modeled using Poisson distribution, but the intensity parameters of the model varied (within any neighborhood) following a Gamma (Γ) distribution. The resulting compound model has the $Var\ [O_{ik}] > \lambda_{ik}$, where overdispersion could be captured and adjusted (Law et al., 2006). Following the work of Law et al. (2006), two Gaussian random-effects terms were included, u_{ik} and s_{ik}, with Equation (5) to construct the targeted Poisson lognormal model. The inclusion of u_{ik} and s_{ik} would help capture the non-spatial and spatial structures in the unknown area-specific relative risks due to unmeasured or latent covariates. Additionally, the s_{ik} term would help adjust for the spatial autocorrelation in the psychotic, non-psychotic and combined mental health datasets, as observed from the global Moran's I test.

So, the Equation (5) becomes,

$$\log\,[\lambda_{ik}] = \log\,[E_{ik}] + \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + u_{ik} + s_{ik} \quad (6)$$

The models given by Equation (6) were fitted using the WinBUGS software (https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/). The prior information for the β_1, β_2, β_3, β_4, β_5 and β_6 terms were specified as a normal distribution with an expected mean of 0 and a precision (1/variance) of 0.00001. For the spatially non-structured (u_{ik}) and structured (s_{ik}) random effect terms, an independent normal distribution, and the intrinsic conditional autoregressive (ICAR) distribution were specified. The prior information of precision parameters for the unknown random effects was specified as a Γ distribution (a,b) with a mean of $\frac{a}{b}$ and variance of $\frac{a}{b^2}$. For this analysis, the prior distribution of Γ (0.001, 0.001) was used for both the random effect terms. The intercept term, β_0 , was assigned with an improper uniform prior, dflat() due to the inclusion of a sum-to-zero constraint on the random effects.

In order to understand the relative contributions of the spatially non-structured (u_{ik}) and structured (s_{ik}) random effect terms, the posterior distribution of the quantity ψ was calculated, which could be expressed as (Arnold, Thomas, Waller, & Conlon, 1999):

$$\psi = \frac{SD_{s_{ik}}}{(SD_{s_{ik}} + SD_{u_{ik}})} \tag{7}$$

where, $SD_{s_{ik}}$ is the empirical marginal standard deviation of s_{ik} and $SD_{u_{ik}}$ is the empirical marginal standard deviation of u_{ik}.

As $\psi \to 1$, the spatially structured random effect (s_{ik}) would dominate the model compared to the non-structured effect (u_{ik}) and so, the variation in the area-specific relative risk due to unmeasured covariates would be mainly spatial in nature. Conversely, when $\psi \to 0$, the non-structured random effect dominates the model and the effect of spatial variation could be considered as negligible.

Initial values were assigned to the parameters, from which the estimation began and converged to the target posterior distribution. The convergence was checked by running two chains with widely differing initial values and by visual inspection of the trace plots, the serial autocorrelation function and the Gelman-Rubin diagnostic. The trace plots were inspected to check whether the samples from the chains scattered around a stable mean, while the autocorrelation graphs were checked to see whether the graphs had approached towards zero. The Gelman-Rubin graphs were checked to observe whether the ratio of the between and within-chain variances converged towards 1.0. Once the convergence had reached, the accuracy of the posterior estimate was assessed using the Monte Carlo (MC) error of the posterior mean for each parameter. The accuracy of the estimation and the number of samples taken to generate the posterior estimate were considered as satisfactory when the MC error was <5% of the sample (posterior) deviation. The deviance information criterion (DIC) and the number of effective parameters in the model (p_D) were recorded for each model to allow the comparison of the models and the selection of the best model.

$$DIC = \overline{D} + p_D \tag{8}$$

where, $\overline{D}$ is the posterior mean of the deviance

The model given by Equation (6) was repeated separately for psychotic, non-psychotic disorders and for each of the vegetation measures (EVI, NDVI, SAVI, Veg_RF and Tree_OD). Hence, a total of 10 models were required for this part of the analyses. The models of the same outcome variable (for example, psychotic or non-psychotic) but using different vegetation measures were compared to understand the differential effect of vegetation measures on the association between vegetation and mental health disorders.

In addition to DIC and p_D, comparisons between the models were made in terms of the area-specific relative risks and the role of different vegetation measures in determining the significance of the association. Based on these comparisons, a suitable vegetation measure that could accurately capture the vegetation coverage in the Toronto area was selected to study the age and sex-stratified association between vegetation and mental health disorders. These age and sex-specific analyses were conducted separately for males and females and for the age groups 0-19, 20-44, 45-60 and 60 above. Hence, a total of 8 models were produced for this part of the analysis. As the rate of substance use was excluded from the age and sex-specific analysis, the models developed could be defined by the Equation (9):

$$\log[\lambda_{ik}] = \log[E_{ik}] + \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_{ik} + s_{ik} \tag{9}$$

3.2.8 Assessment of the relative risk of mental health disorders due to the variations in vegetation content

The relative risk values from the models of the five vegetation measures were explored and checked if they substantially differed from each other. The posterior mean values from the Bayesian models and the median and the interquartile ranges of the relative risk values were assessed using box plot diagrams to observe the differences in absolute magnitude. Afterward, the results from the Bayesian spatial modeling (95% CI, DIC and p_D) and the risk value assessments were used to select one (out of the five) vegetation measure to map the relative risk of different mental health disorders in the Toronto area.

The model associated with this vegetation measure was then used to map the relative risks of psychotic, non-psychotic and combined mental health disorders in the study area. The relative risk being mapped was owing to the variations in vegetation content after adjusting for potential confounders and unmeasured covariates. Equations (3) and (6) show that the relative risk can be defined using the following model components:

$$r_{ik} = \exp[\beta_0] * \exp[\beta_1 X_{1i}] * \exp[\beta_2 X_{2i}] * \exp[\beta_3 X_{3i}] * \exp[\beta_4 X_{4i}] * \exp[\beta_5 X_{5i}] * \exp[\beta_6 X_{6i}] * \exp[u_{ik}] * \exp[s_{ik}] \quad (10)$$

As the rate of substance use was excluded from the age and sex-specific analysis of this study, the relative risk of combined mental health disorders for males and females and for each of the age-groups is defined by:

$$r_{ik} = \exp[\beta_0] * \exp[\beta_1 X_{1i}] * \exp[\beta_2 X_{2i}] * \exp[\beta_3 X_{3i}] * \exp[\beta_4 X_{4i}] * \exp[\beta_5 X_{5i}] * \exp[u_{ik}] * \exp[s_{ik}] \quad (11)$$

The details of the methodology from the data preparation to the BSM are summarized in *Appendix E.*

Chapter 4 Results

4.1 Vegetation Indices

Figure 3 illustrates the false-color composite of the raw Landsat-8 image, the three vegetation indices (EVI, NDVI and SAVI) and the area-based measures of vegetation cover (Veg_RF and Tree_OD). The false-color composite image displayed here utilizes the traditional color infrared image visualization technique for satellite images and the band combination of near-infrared, red and green (instead of red, green and blue) to illustrate vegetation in bright red color vibrantly (Jensen, 2009). Accuracy assessments revealed quite high accuracies of the final LULC model used to derive the Veg_RF variable. The user's accuracy and the Kappa coefficient values for the final LULC model were 0.967 and 0.909, respectively. The developed LULC model suggested that 22.5% of Toronto was covered by vegetation in 2016.

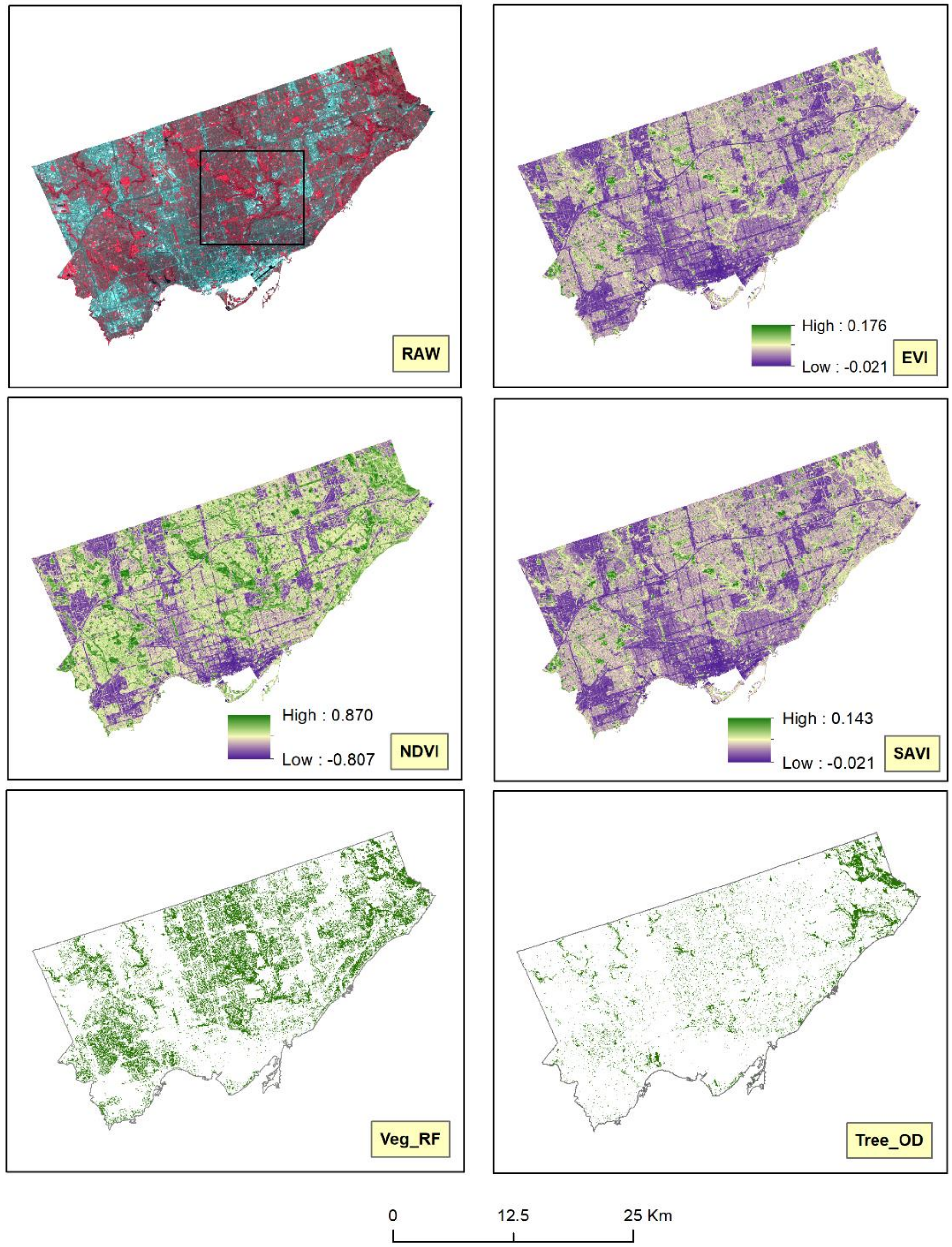


Figure 3: The study area showing the macro-scale differences between the three vegetation indices (EVI, NDVI and SAVI) and the area-based measures of vegetation cover (Veg_RF and Tree_OD). The shades of green represent the vegetation-covered areas for all the three vegetation indices and the solid green color represents vegetation cover in the area-based measures. The black selection box in the raw image represents a portion of the study area that was zoomed in Figure 4 for better visualization of the micro-scale differences.

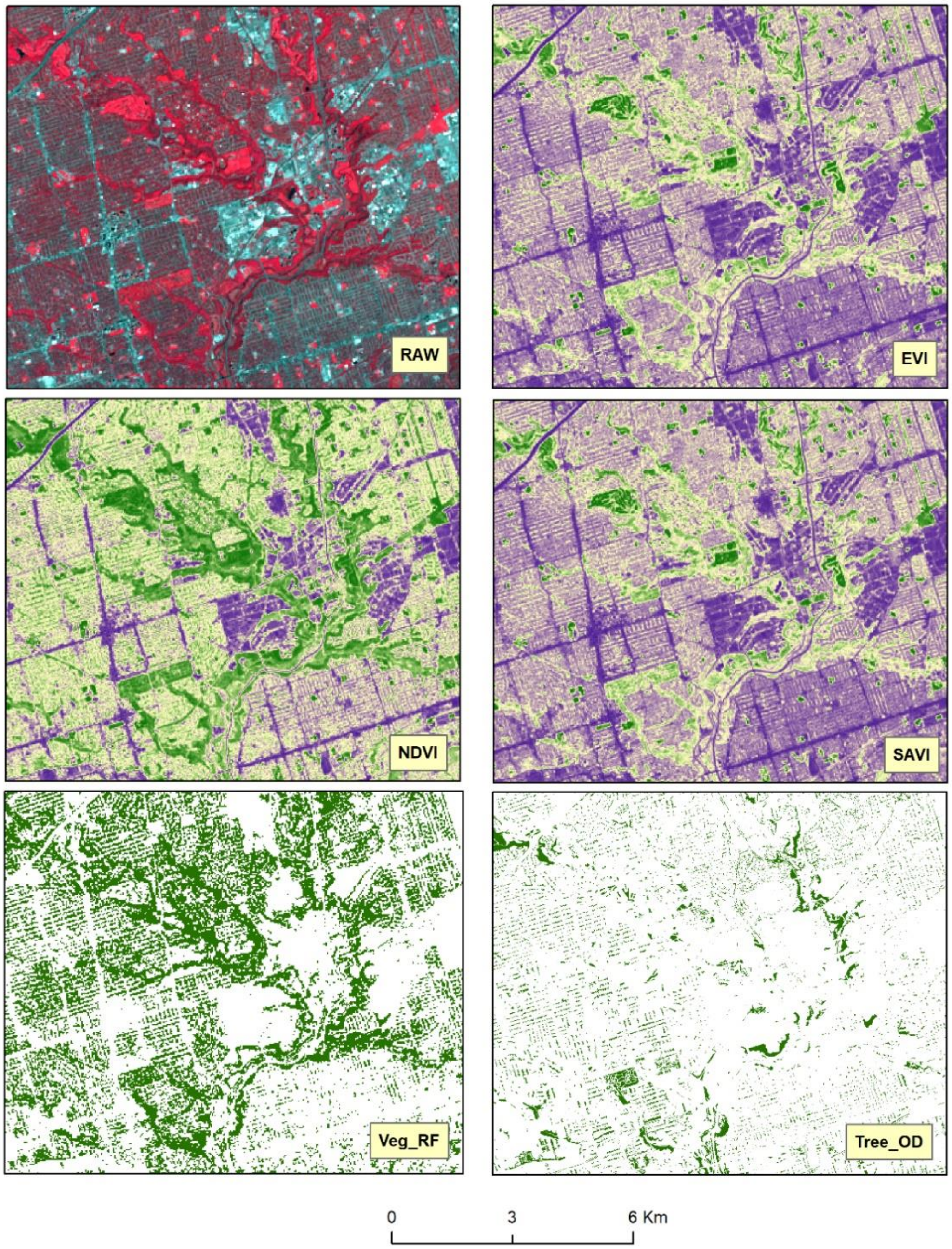
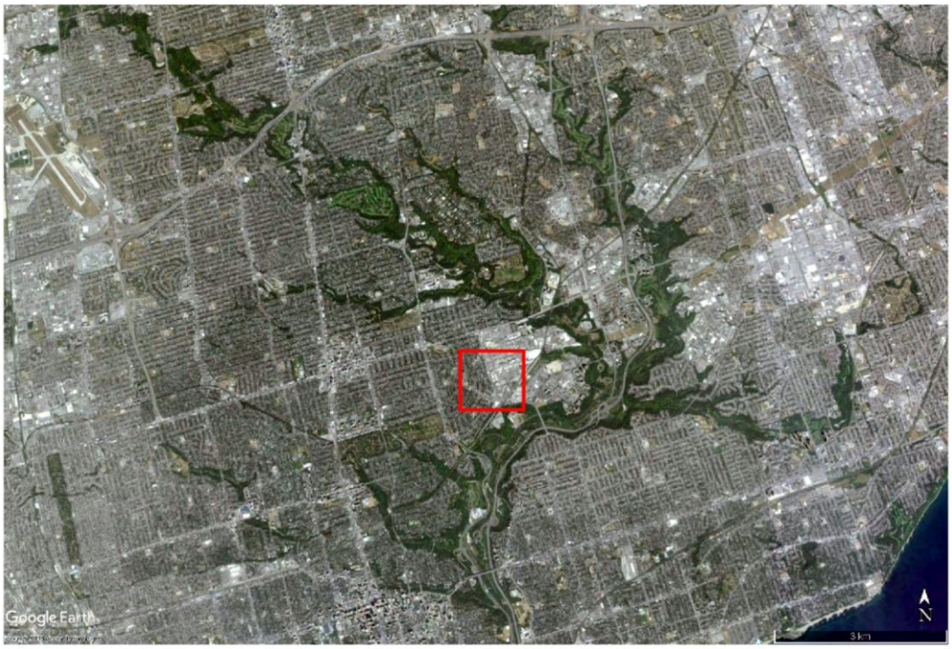


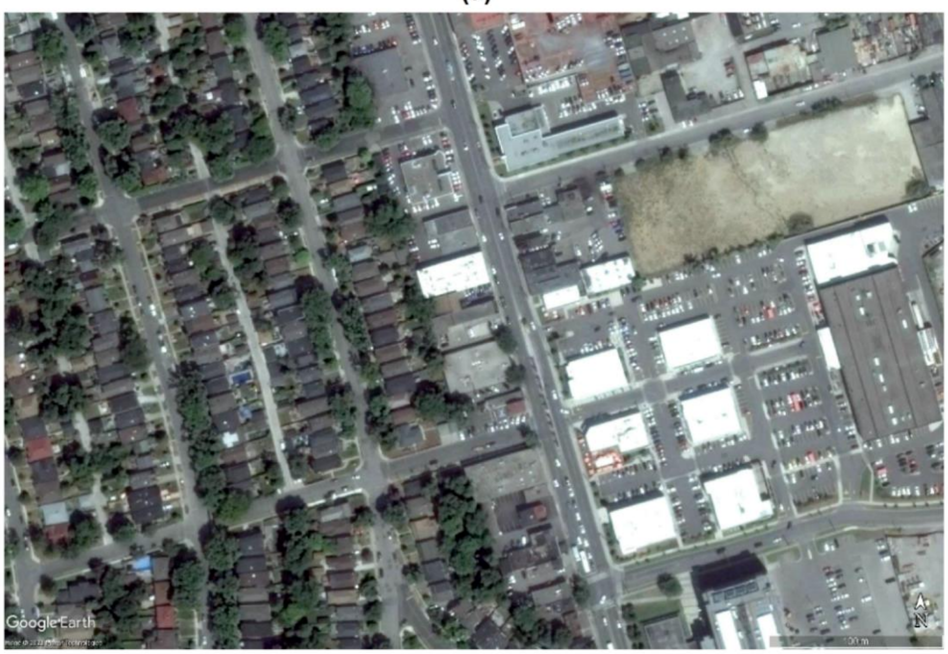
Figure 4: A portion of the study area showing the micro-scale differences between the three vegetation indices (EVI, NDVI and SAVI) and the area-based measures of vegetation cover (Veg_RF and Tree_OD). The shades of green represent the vegetation-covered areas for all the three vegetation indices, with darker shades of green representing dense and healthy vegetation. The yellow and the purple areas mainly represent the non-vegetation areas in the indices. The solid green color represents vegetation cover, while the white color represents non-vegetation cover in the area-based measures.

Comparing the different vegetation measures could shed further insights into their differences. Figures 3 and 4 show that the three constructed vegetation indices showed a gradation of green color to illustrate both the density and health of the vegetation cover. Figure 4 also shows that the yellow patches in the NDVI image contained a marked presence of green color compared to the other two indices. On closer inspection and further magnification of Figure 4, the Google Earth images in Figure 5 indicate that these yellow and small green patches actually represented the built-up structures with surrounding vegetation, respectively. Hence, there is evidence of spectral confusion or falsely detecting other non-vegetation features as vegetation-covered areas.

Interestingly, despite having different computational processes (Table 2), EVI and SAVI could be seen as more similar to each other compared to NDVI. In contrast, both the area-based measures of vegetation only showed the areal-extent of vegetation, as indicated by the solid green color. The Tree_OD data had severely underestimated the vegetation content compared to the other four satellite-derived vegetation measures. The areal extent of vegetation covers detected by Veg_RF matched more with NDVI, compared to EVI or SAVI.



(a)



(b)

Figure 5: Google Earth images showing (a) a segment of the study area with vegetation cover and (b) a magnified image of the segment

4.2 The association between vegetation and mental health disorders

4.2.1 The association between different measures of vegetation and psychotic and non-psychotic disorders

The results of the Bayesian spatial modeling were used to analyze the association between various measures of vegetation and psychotic and non-psychotic disorders. The results of the analyses are tabulated in Table 5. The results indicate that only EVI and SAVI were significantly associated with both psychotic and non-psychotic disorders. These two vegetation indices were negatively associated with the number of psychotic and non-psychotic disorder cases, implying that an increase in EVI and SAVI values could decrease the number of mental health disorder cases in the study area. The magnitude of the association between EVI and psychotic disorders was $\boldsymbol{\beta_1}$= -4.056 (95% CI: -8.147, -0.025) and that between EVI and non-psychotic disorders was $\boldsymbol{\beta_1}$= -2.442 (95% CI: -4.735, -0.172). Similarly, the magnitude of the association of SAVI with psychotic disorders was $\boldsymbol{\beta_1}$= -3.676 (95% CI: -7.350, -0.008) and with that of non-psychotic disorders was $\boldsymbol{\beta_1}$= -2.213 (95% CI: -4.372, -0.121). Neither NDVI nor any of the area-based vegetation measures (Veg_RF and Tree_OD) have shown any significant association with the psychotic and non-psychotic disorder cases.

Amongst the confounding variables, ethnic concentration (β_3), residential instability (β_4) and the rate of substance use disorder (β_6) have shown statistically significant associations with both the psychotic and non-psychotic disorder cases. However, only ethnic concentration has shown a negative association, implying that an increase in ethnic concentration may lead to a decrease in the psychotic and non-psychotic disorder cases in the study area. In contrast, material deprivation (β_2) was found to be significantly and positively associated with psychotic disorders. The dependency (β_5) variable did not exhibit any significant association with any of the two outcome variables.

The values of $\boldsymbol{\psi}$ are greater than 0.50 in all the ten models and are all statistically significant. The $\boldsymbol{\psi}$ values for the models of psychotic disorders are close to 0.50 and therefore, the spatially structured random effect term (s_{ik}) and the non-structured random effect term (u_{ik}) are almost equally dominant in the models. However, the values of $\boldsymbol{\psi}$ in the models for non-psychotic disorders are greater than 0.70 and are closer to 1 ($\boldsymbol{\psi} \rightarrow 1$), showing that s_{ik} had dominated each of the models compared to u_{ik}. Therefore, the variations in the area-specific relative risk due to unmeasured covariates in the study area had notable spatial structures for both the psychotic and non-psychotic disorder cases.

No discernible differences in the values of DIC and the number of effective parameters (p_D) are evident for the models analyzing the association between vegetation and psychotic disorders. Similar results were obtained for the models on non-psychotic disorders. These results demonstrated the fact that for a specific outcome variable (for example, psychotic or non-psychotic disorders) using different vegetation measures did not have any effect on the goodness of fit and the model parsimony. The most notable change observed from the results, therefore, is the difference in the significance of the association with the vegetation variables. The findings suggest that a significant association is detected only with the vegetation indices, more specifically, with the EVI and SAVI.

Table 5: Summaries of results from Bayesian spatial modeling to analyze the association between vegetation and psychotic and non-psychotic disorders. The italicized values are significant at a 95% credible interval (CI)

Posterior means summaries	**EVI**	**NDVI**	**SAVI**	**Veg_RF**	**Tree_OD**
Psychotic disorders					
$\boldsymbol{\beta_0}$	*-0.287*	-0.148	*-0.286*	*-0.477*	*-0.492*
(95% CI)	*(-0.514, -0.057)*	(-0.508, 0.206)	*(-0.513, -0.059)*	*(-0.583, -0.375)*	*(-0.591, -0.395)*
$\boldsymbol{\beta_1}$	*-4.056*	-0.626	*-3.676*	-0.001	-0.001
(95% CI)	*(-8.147, -0.025)*	(-1.249, 0.000)	*(-7.350, -0.008)*	(-0.005, 0.004)	(-0.006, 0.005)
$\boldsymbol{\beta_2}$	*0.122*	*0.117*	*0.121*	*0.108*	*0.112*
(95% CI)	*(0.077, 0.166)*	*(0.073, 0.161)*	*(0.076, 0.165)*	*(0.062, 0.153)*	*(0.068, 0.156)*
$\boldsymbol{\beta_3}$	*-0.118*	*-0.118*	*-0.117*	*-0.121*	*-0.123*
(95% CI)	*(-0.169, -0.064)*	*(-0.169, -0.065)*	*(-0.169, -0.063)*	*(-0.172, -0.067)*	*(-0.175, -0.067)*
$\boldsymbol{\beta_4}$	*0.179*	*0.180*	*0.179*	*0.179*	*0.181*
(95% CI)	*(0.135, 0.221)*	*(0.137, 0.221)*	*(0.135, 0.221)*	*(0.136, 0.221)*	*(0.138, 0.223)*
$\boldsymbol{\beta_5}$	-0.057	-0.057	-0.056	-0.057	-0.061
(95% CI)	(-0.124, 0.011)	(-0.125, 0.011)	(-0.124, 0.012)	(-0.126, 0.012)	(-0.130, 0.008)
$\boldsymbol{\beta_6}$	*0.041*	*0.041*	*0.041*	*0.041*	*0.041*
(95% CI)	*(0.033, 0.049)*	*(0.033, 0.049)*	*(0.033, 0.049)*	*(0.033, 0.049)*	*(0.033, 0.049)*
$\boldsymbol{\psi}$	*0.537*	*0.519*	*0.539*	*0.501*	*0.522*
(95% CI)	*(0.231, 0.792)*	*(0.223, 0.779)*	*(0.236, 0.792)*	*(0.203, 0.787)*	*(0.213, 0.798)*
$\mathbf{p}_{\boldsymbol{D}}$	102.66	102.589	102.642	103.662	103.683
DIC	1271.530	1271.580	1271.560	1272.160	1272.110
Non-psychotic disorders					
$\boldsymbol{\beta_0}$	0.098	0.015	0.098	*-0.073*	*-0.062*
(95% CI)	(-0.031, 0.230)	(-0.195, 0.227)	(-0.037, 0.236)	*(-0.135, -0.012)*	*(-0.122, -0.003)*
$\boldsymbol{\beta_1}$	*-2.442*	-0.081	*-2.213*	0.002	0.004
(95% CI)	*(-4.735, -0.172)*	(-0.446, 0.280)	*(-4.372, -0.121)*	(-0.002, 0.006)	(-0.001, 0.008)
$\boldsymbol{\beta_2}$	0.014	0.009	0.013	0.015	0.007
(95% CI)	(-0.014, 0.041)	(-0.019, 0.036)	(-0.015, 0.040)	(-0.012, 0.041)	(-0.020, 0.033)
$\boldsymbol{\beta_3}$	*-0.114*	*-0.115*	*-0.114*	*-0.115*	*-0.107*
(95% CI)	*(-0.147, -0.082)*	*(-0.148, -0.082)*	*(-0.146, -0.081)*	*(-0.147, -0.083)*	*(-0.140, -0.075)*
$\boldsymbol{\beta_4}$	*0.055*	*0.057*	*0.055*	*0.062*	*0.056*
(95% CI)	*(0.028, 0.082)*	*(0.029, 0.084)*	*(0.028, 0.082)*	*(0.035, 0.089)*	*(0.029, 0.082)*
$\boldsymbol{\beta_5}$	0.007	0.006	0.007	-0.002	0.007
(95% CI)	(-0.032, 0.046)	(-0.034, 0.045)	(-0.031, 0.046)	(-0.041, 0.037)	(-0.031, 0.046)
$\boldsymbol{\beta_6}$	*0.011*	*0.011*	*0.011*	*0.011*	*0.011*
(95% CI)	*(0.005, 0.017)*	*(0.005, 0.017)*	*(0.005, 0.017)*	*(0.005, 0.017)*	*(0.006, 0.017)*
$\boldsymbol{\psi}$	*0.750*	*0.754*	*0.750*	*0.744*	*0.755*
(95% CI)	*(0.595, 0.863)*	*(0.595, 0.867)*	*(0.596, 0.863)*	*(0.593, 0.860)*	*(0.601, 0.866)*
$\mathbf{p}_{\boldsymbol{D}}$	126.554	127.088	126.678	125.982	126.780
DIC	1591.070	1591.540	1591.290	1590.810	1590.750

The relative risk values (r_{ik}) of psychotic and non-psychotic disorders, as defined by Equation (10), for each of the vegetation measures are shown in Figure 6. The median and the interquartile range of the box plots show that there are substantial differences in the relative risks for the psychotic and non-psychotic disorders. However, the relative risk values are very similar for the five vegetation measures in both these mental health disorder categories.

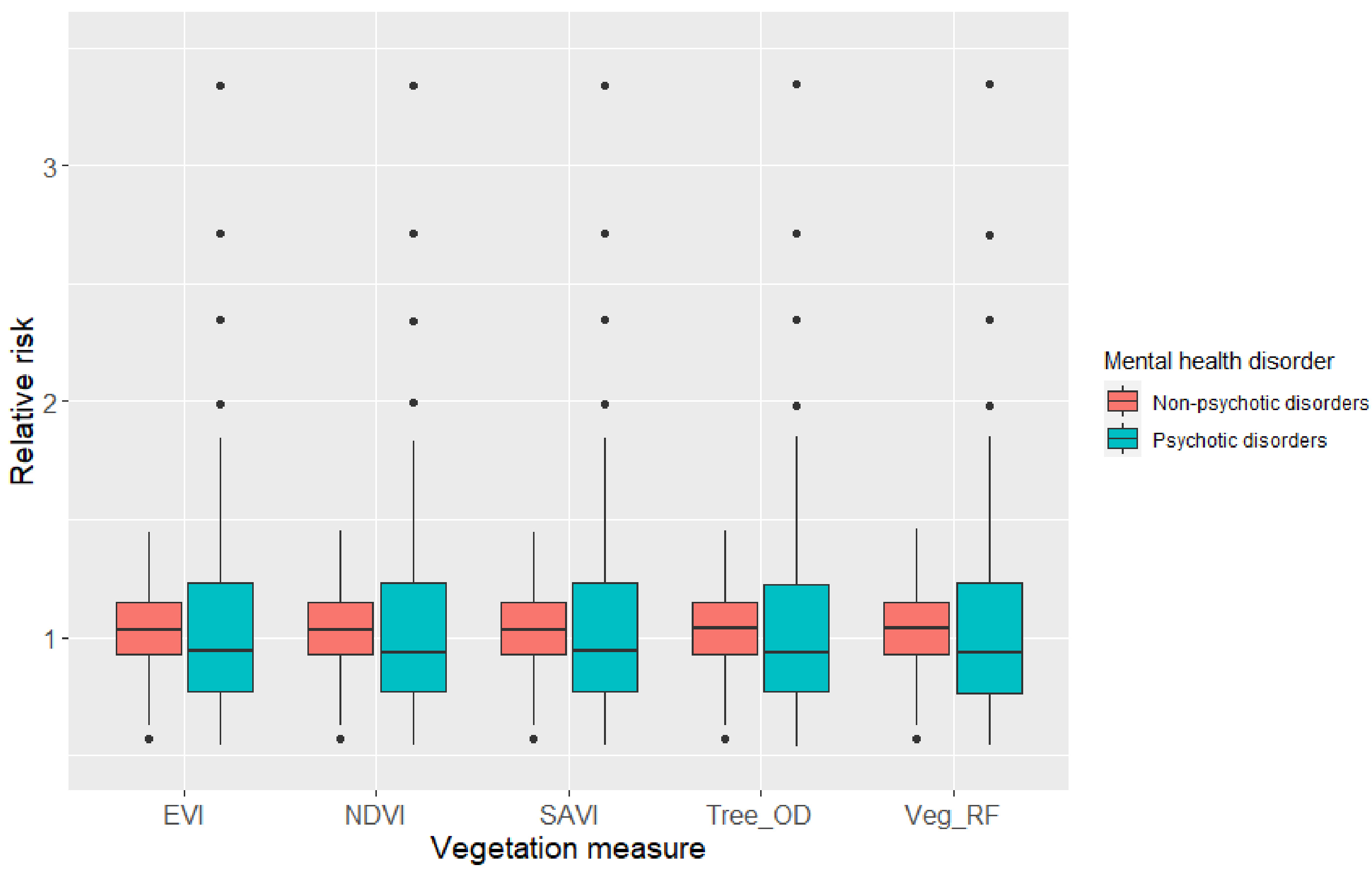


Figure 6: Box plot diagram showing the posterior mean of the relative risks of psychotic and non-psychotic disorders for the five different vegetation measures in the 140 neighborhoods in Toronto.

4.2.2 The association between EVI and the age and sex-stratified mental health disorder cases

As the assessment of DIC and p_D did not aid in the selection of a particular model over another (and hence, a particular vegetation measure), only the vegetation measures that had shown significant associations with both the psychotic and non-psychotic disorders were considered for the age and sex-specific analyses. Therefore, only EVI and SAVI could be considered as suitable candidates for this part of the study.

However, Table 5 and Figure 6 suggest that the EVI and SAVI had functioned in a similar manner while modeling the association between vegetation and mental health disorders. Both these indices showed significant associations with the two types of disorders, had a similar magnitude of association and DIC and p_D values. Therefore, the selection between these two indices was made based on the computational differences between the two indices, which may cause one of these indices to perform comparatively better in an urban setting. In this regard, the formulas for EVI and SAVI in Table 2 were consulted. The formulas indicate that the EVI had undergone three specific corrections as opposed to only one for SAVI. For example, the EVI was corrected for the atmospheric disturbances using two separate constants (C1 and C2 in Table 2) and also for the canopy background cover (L_{EVI} in Table 2). In contrast, SAVI was only corrected for the soil brightness factor (L_{SAVI} in Table 2). Consequently, EVI was chosen as the vegetation measure to analyze the association between vegetation and mental health disorders in males and females from different age groups. A detailed explanation of how these corrections could create a major difference in the detection of vegetation cover in an urban setting is provided in Section 5 (Discussions).

The results of the association between vegetation and the age and sex-specific mental health disorders are tabulated in Table 6. The results suggest that the mental health conditions of males, from age groups 0-19 and 20-44, are significantly affected by the presence of vegetation cover. Contrastingly, only females from the age group 20-44 are influenced by urban greenery (represented by EVI). The magnitude of the association between vegetation and mental health disorders for males was $\boldsymbol{\beta_1}$= -7.009 (95% CI: -13.130, -0.980) and $\boldsymbol{\beta_1}$= -4.544 (95% CI: -8.224, -0.895) for the age groups 0-19 and 20-44, respectively. This magnitude of the association for females in the age group 20-44 was $\boldsymbol{\beta_1}$= -3.513 (-6.289, -0.681). Therefore, the results suggest that increased vegetation cover could have an ameliorating effect on both young males and females. In particular, males from the age group 0-19 could be most benefitted due to the presence of vegetation.

The socioeconomic covariates demonstrated varying degrees of association with mental health disorder cases. The ethnic concentration was negatively and significantly associated with the mental health disorder cases of both the sexes and for all the age groups. Interestingly, dependency was found to be negatively associated with the mental health disorder cases of both the sexes and for age groups 20-44 and 45-64. The material deprivation and residential instability showed significant and positive associations with mental health disorder cases of males and females in the age groups 20-44 and 45-64. Additionally, the material deprivation was significant for males aged 65 years and above.

The $\boldsymbol{\psi}$ values for all the models in this part of the study are significant and greater than 0.70. Since all the $\boldsymbol{\psi}$ values are closer to 1 ($\psi \rightarrow 1$), the variations in the area-specific relative risk of mental health disorders of both sexes and for all age groups were mainly influenced by the unmeasured spatial covariates.

Table 6: Summaries of results from Bayesian spatial modeling to analyze the associations between EVI and combined mental health disorders for males and females of different age groups. The italicized values are significant at a 95% credible interval (CI)

Posterior means summaries	**0-19**	**20-44**	**45-64**	**65+**
Males				
$\boldsymbol{\beta_0}$ (95% CI)	*0.550 (0.232, 0.873)*	*0.274 (0.078, 0.472)*	0.138 (-0.056, 0.333)	*0.221 (0.004, 0.438)*
$\boldsymbol{\beta_1}$ (95% CI)	*-7.009 (-13.130, -0.980)*	*-4.544 (-8.224, -0.895)*	-2.920 (-6.585, 0.721)	-3.841 (-7.913, 0.166)
$\boldsymbol{\beta_2}$ (95% CI)	-0.035 (-0.097, 0.027)	*0.095 (0.056, 0.134)*	*0.132 (0.093, 0.169)*	*0.053 (0.012, 0.093)*
$\boldsymbol{\beta_3}$ (95% CI)	*-0.192 (-0.265, -0.121)*	*-0.148 (-0.194, -0.102)*	*-0.156 (-0.201, -0.110)*	*-0.079 (-0.127, -0.030)*
$\boldsymbol{\beta_4}$ (95% CI)	-0.013 (-0.073, 0.046)	*0.080 (0.040, 0.119)*	*0.136 (0.098, 0.174)*	*0.075 (0.035, 0.116)*
$\boldsymbol{\beta_5}$ (95% CI)	*-0.131 (-0.228, -0.034)*	*-0.074 (-0.137, -0.012)*	*-0.109 (-0.174, -0.044)*	-0.022 (-0.086, 0.043)
$\boldsymbol{\psi}$ (95% CI)	*0.716 (0.520, 0.899)*	*0.795 (0.647, 0.904)*	*0.742 (0.558, 0.886)*	*0.817 (0.688, 0.908)*
$\mathbf{p}_{\boldsymbol{D}}$	96.521	109.930	107.499	80.095
DIC	1134.110	1329.180	1323.320	1184.860
Females				
$\boldsymbol{\beta_0}$ (95% CI)	*0.325 (0.001, 0.645)*	*0.292 (0.141, 0.439)*	0.137 (-0.017, 0.292)	*0.214 (0.015, 0.413)*
$\boldsymbol{\beta_1}$ (95% CI)	-1.624 (-7.632, 4.448)	*-3.513 (-6.289, -0.681)*	-1.376 (-4.274, 1.533)	-2.934 (-6.660, 0.784)
$\boldsymbol{\beta_2}$ (95% CI)	-0.029 (-0.089, 0.032)	*0.096 (0.066, 0.125)*	*0.093 (0.063, 0.123)*	0.023 (-0.016, 0.061)
$\boldsymbol{\beta_3}$ (95% CI)	*-0.262 (-0.334, -0.192)*	*-0.186 (-0.221, -0.151)*	*-0.144 (-0.180, -0.108)*	*-0.104 (-0.150, -0.058)*
$\boldsymbol{\beta_4}$ (95% CI)	0.013 (-0.046, 0.072)	*0.048 (0.019, 0.078)*	*0.071 (0.041, 0.101)*	*0.068 (0.030, 0.107)*
$\boldsymbol{\beta_5}$ (95% CI)	-0.069 (-0.165, 0.026)	*-0.068 (-0.115, -0.021)*	*-0.064 (-0.112, -0.016)*	0.034 (-0.025, 0.092)
$\boldsymbol{\psi}$ (95% CI)	*0.738 (0.544, 0.906)*	*0.799 (0.675, 0.892)*	*0.779 (0.632, 0.887)*	*0.837 (0.729, 0.914)*
$\mathbf{p}_{\boldsymbol{D}}$	95.835	102.474	100.769	93.684
DIC	1142.020	1368.690	1358.880	1258.990

The relative risks of mental health disorders for males and females in the age groups 0-19, 20-44, 45-64 and 65+ in the neighborhoods of Toronto are compared in Figure 7. The relative risks were calculated using Equation (11). The results indicate that the relative risks are almost similar for the males and females of ages 0-19 and 65+. However, for the age groups 20-44 and 45-64, females have higher relative risks than males. The median values for the relative risks of both the sexes in all the four different age groups are greater than 1, showing elevated risks of developing mental health disorders. The risks are particularly high for females in the age groups 20-44 and 45-64, whereas, for males, the vulnerable age groups are 20-44 and 65+.

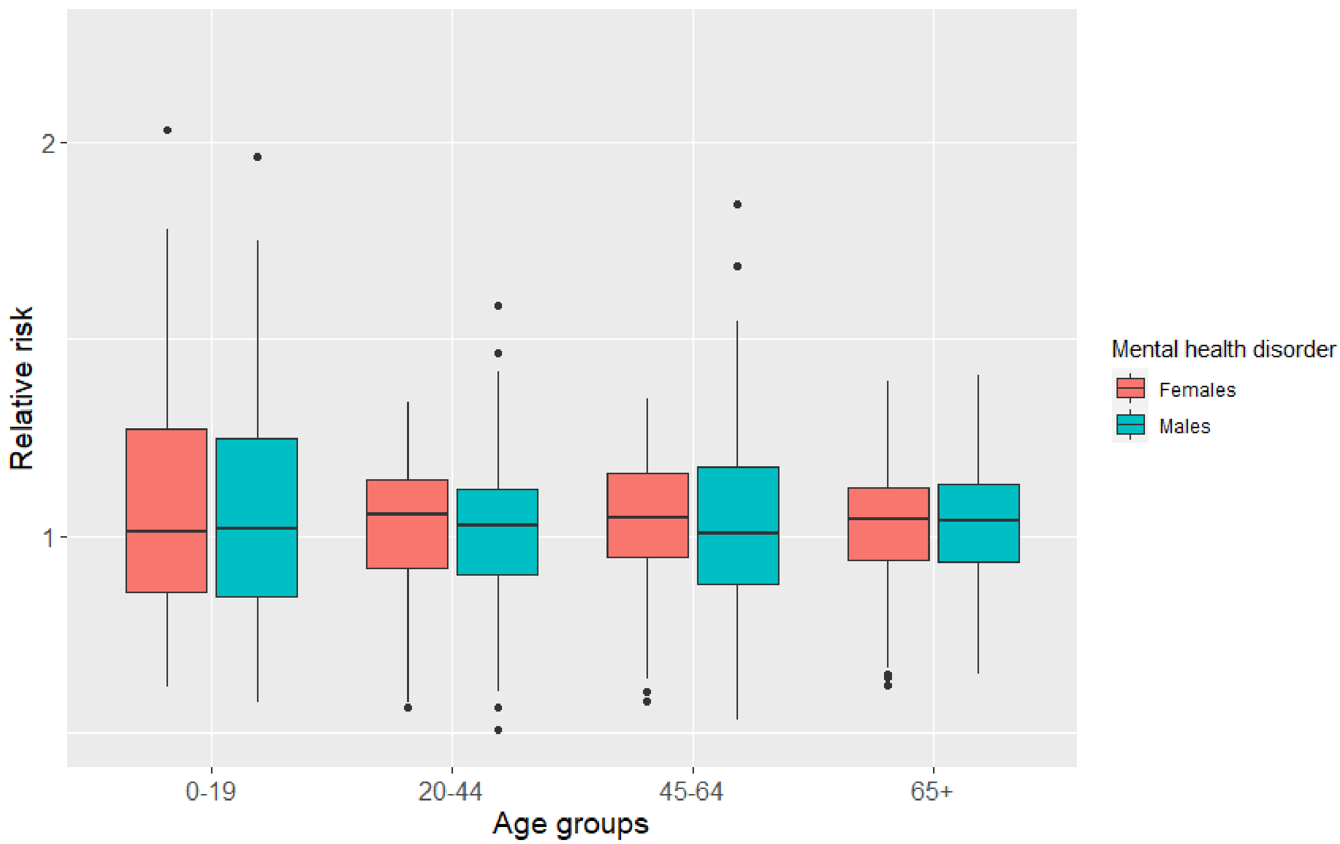


Figure 7: Box plot diagram showing the posterior mean of the relative risks of combined mental health disorders for males and females in the age groups 0-19, 20-44, 45-64 and 65+ in the 140 neighborhoods in Toronto.

4.2.3 The spatial distribution of the relative risk of psychotic, non-psychotic and combined mental health disorders

Contrary to the non-spatial depiction of the relative risks using box-plots, histograms and other different forms of charts, illustrating relative risks using maps can help accurately identify the high-risk areas. The relative risk (r_{ik}) from the EVI models for the psychotic and non-psychotic disorders are shown in Figure 8(a) and (b), respectively. The areas with relative risk values > 1 could be interpreted as areas with high risks from psychotic or non-psychotic disorders due to reduced vegetation cover after adjusting for the risks from material deprivation, ethnic concentration, residential instability, dependence, substance use disorders and the unmeasured covariates.

Figure 8(a) shows that neighborhoods with the relative risk of psychotic disorders > 1 were mostly clustered in the southern part and extended from the west to east. There were six neighborhoods with very high risk (r_{ik} > 1.75) in the southcentral part of Toronto. In contrast, Figure 8(b) reveals that the neighborhoods with the relative risk of non-psychotic disorders > 1 cover much of the southern and the northcentral parts of Toronto. When Figure 8(a) and 8(b) are considered together, it could be observed that the neighborhoods with high risk (r_{ik} > 1) of psychotic disorders were also at high risk from non-psychotic disorders. However, unlike the relative risk for psychotic disorders, the relative risk from non-psychotic disorders did not exhibit very high values and was mostly below the value of 1.5. These two relative risk maps suggest that the northern part of Toronto is relatively at less risk compared to the southern part.

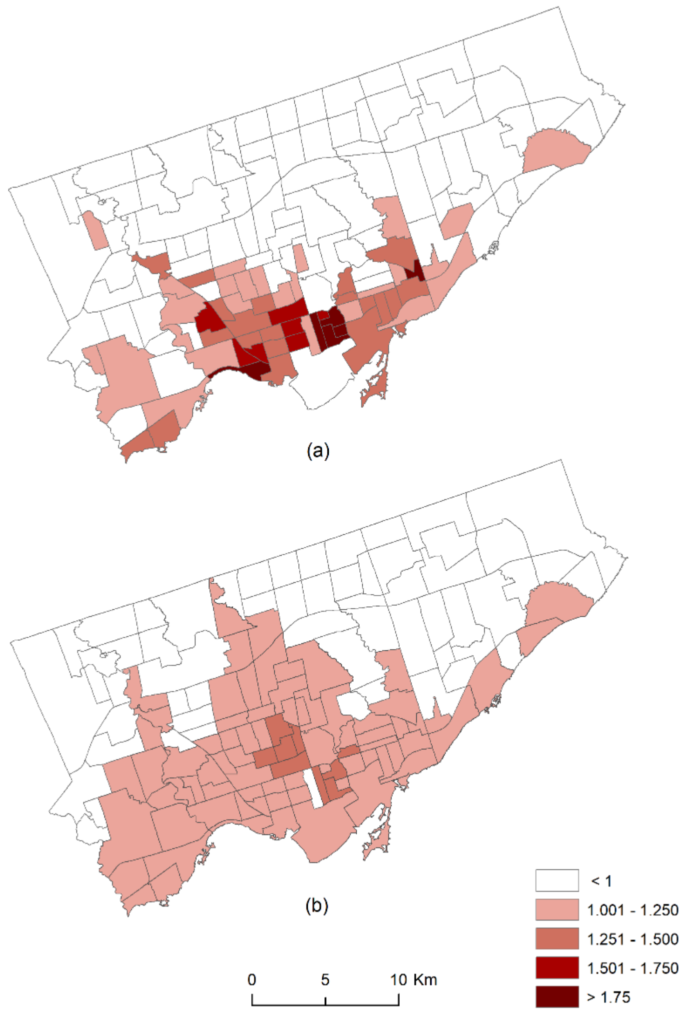


Figure 8: The posterior mean of the relative risk (r_{ik}) of (a) psychotic and (b) non-psychotic disorders.

Figures 9 and 10 show the relative risk of combined mental health disorders due to variations in vegetation cover, after adjusting for the risks from material deprivation, ethnic concentration, residential instability, dependence and the unmeasured covariates. The relative risk maps for males and females are similar for all age groups, suggesting that the spatial distribution of relative risk for mental health disorders is nearly identical for both these sexes. However, interesting variations in the distributions of relative risks could be observed by comparing the maps of different age groups for each of the sexes.

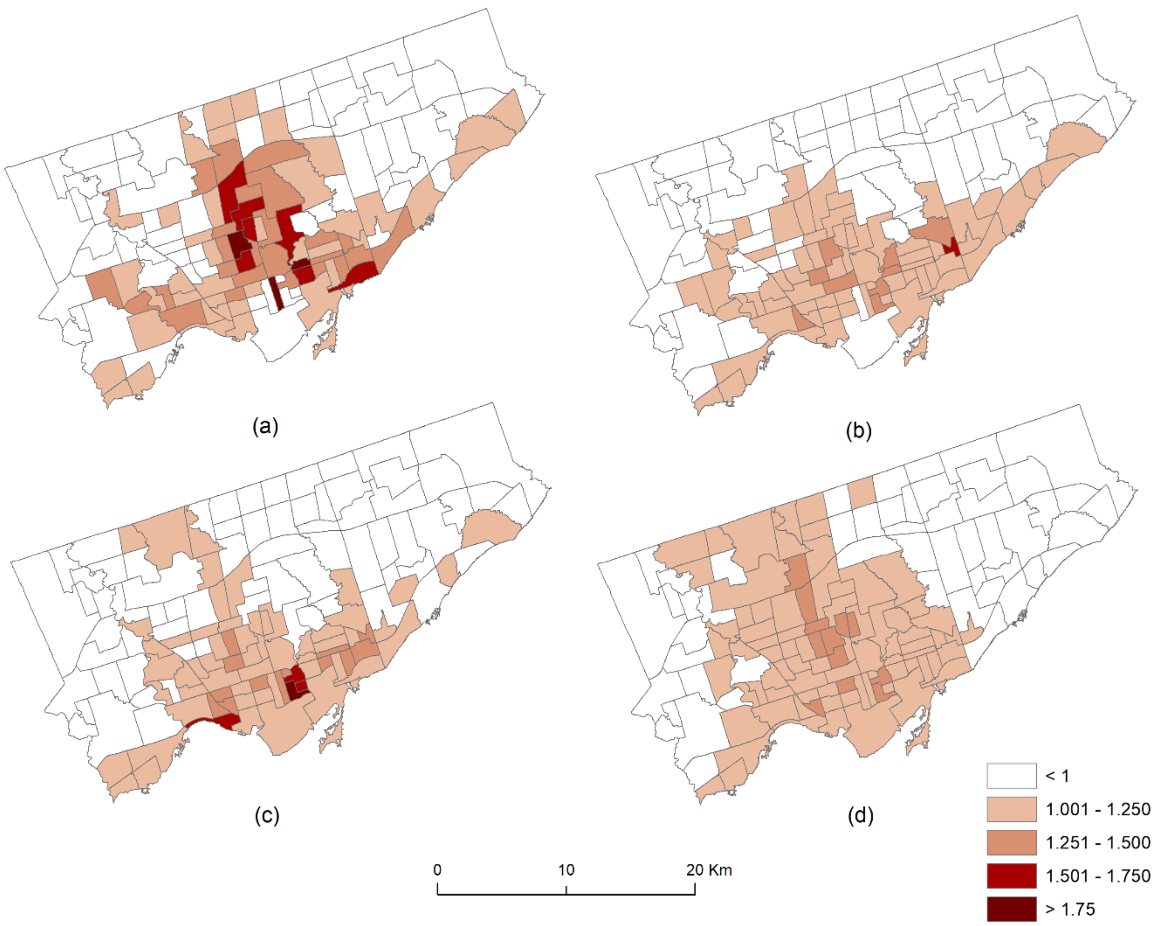


Figure 9: The posterior mean of the relative risk (r_{ik}) of males for the age-groups (a) 0-19, (b) 20-44, (c) 45-64 and (d) 65+

The maps for males and females from age-group 0-19 show that the high-risk neighborhoods ($r_{ik} > 1$) are located in the central part of the Toronto area. In contrast, the high-risk areas for the age-group 20-44 for both the sexes are located in the southern parts of the study area. Although the high-risk neighborhoods for the age-group 45-64, for both males and females, are also mostly located in the southern portion, Figure 10(c) suggests that the risk for females in this age group is distributed over a larger area than the males. A good portion of the central-western part of Toronto is at high risk from mental health disorders of females belonging to the age group 45-64 years. However, for the people (both sexes) in the age group 65+, the neighborhoods with a high relative risk of developing mental health disorders extend from south to northward direction.

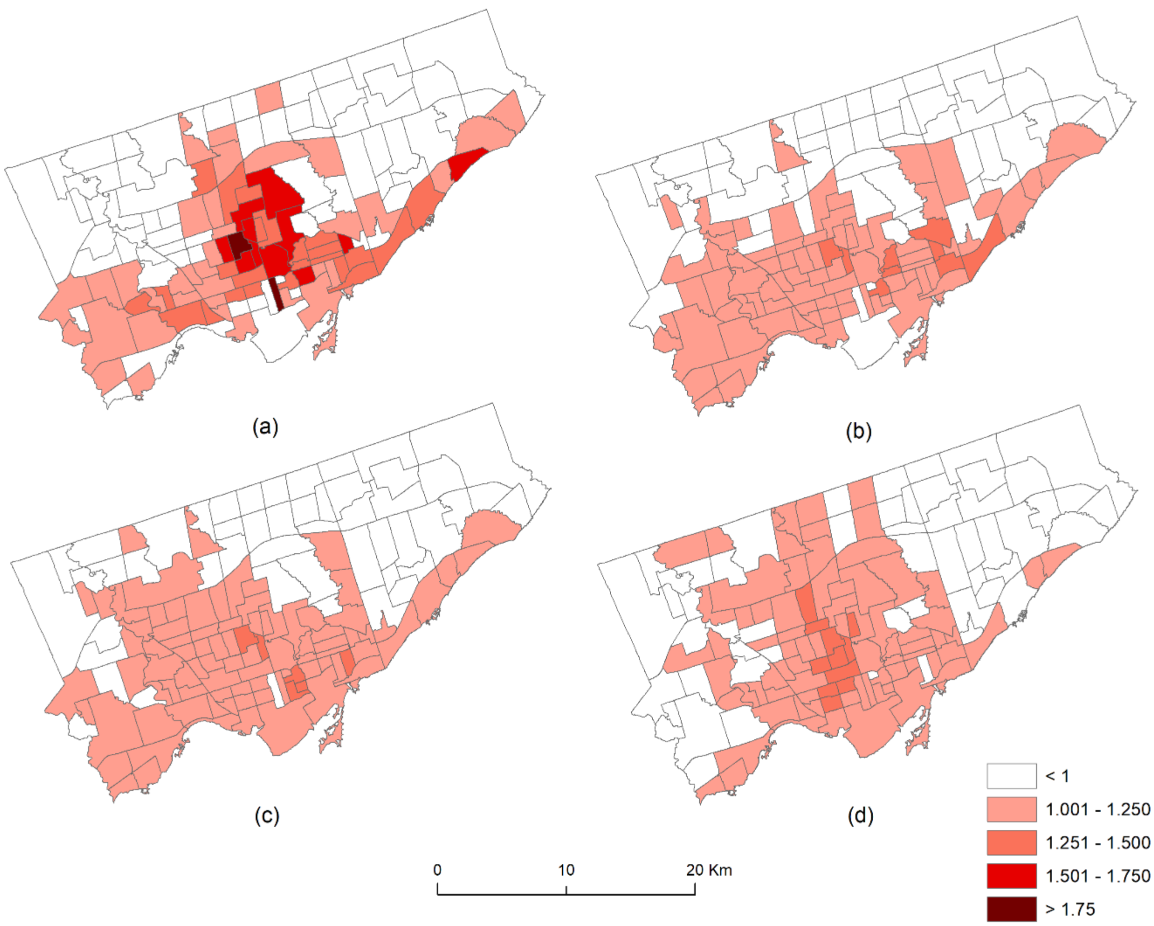


Figure 10: The posterior mean of the relative risk (r_{ik}) of females for the age-groups (a) 0-19, (b) 20-44, (c) 45-64 and (d) 65+

Chapter 5 Discussions

Based on the knowledge from available literature, this is the first study that employed Bayesian spatial statistics to elucidate the performances of different vegetation measures in identifying a significant association between vegetation and mental health. This study provided empirical evidence that the type of vegetation measure in the model could heavily influence the significance of the association with mental health. Furthermore, the log-linear models (specifically, the $\boldsymbol{\psi}$ values) revealed a strong dominance of the spatially structured unmeasured and latent covariates during the relative risk estimations. These latent covariates, if not adjusted in an epidemiological study, could potentially affect the detection of a significant association between vegetation and mental health and can also bias the risk estimation. Results strongly suggest that the satellite-based vegetation indices, which are corrected for atmospheric disturbances, canopy background noise and soil brightness, could help detect a significant association between vegetation and different types of mental health disorders. This could be due to the ability of these indices to provide detailed information on both the areal extent and the health of the surrounding vegetation and thus, capturing people's true exposure to surrounding greenness. Age and sex-specific analyses suggest that the young people, particularly males from the age-group 0-19 and both males and females from the age group 20-44, could be highly susceptible to reduced vegetation cover. For the older adults, from age groups 45-64 and 65+, the socioeconomic factors are more significantly influential than the variations in vegetation cover. Mapping the relative risks of mental health disorders for individual age-groups revealed both micro and macro scale variations in the spatial distribution of the mental health disorder risks, which could provide valuable information for key and targeted intervention strategies.

5.1 The differential impact of vegetation measures on the association between vegetation and mental health disorders

This study found that the area-based measures of vegetation cover (Veg_RF and Tree_OD) could not capture any significant association between vegetation and the psychotic and non-psychotic disorders. This difference could be explained in terms of the differences in their functionality. Every day people are regularly exposed to different forms of vegetation in their surroundings (Markevych et al., 2017; Takayama et al., 2017), which several studies attempted to characterize using the term "surrounding greenness." These studies found that both the density and health of vegetation are vital components for measuring the surrounding greenness in an area (Bezold et al., 2018; James, Banay, Hart, & Laden, 2015). The extent to which vegetation cover can impart mental health benefits is directly dependent on the intensity and quality of the exposure to surrounding greenness, which in turn, depends on the richness of the vegetation cover and the duration of exposure (Dzhambov et al., 2018b; Markevych et al., 2017). In this regard, the area-based vegetation measures were simply based on the percentage of vegetation or tree cover in a neighborhood. Therefore, the values could not vary by the level of surrounding greenness to which people were exposed. Consequently, the association being analyzed, using these area-based measures, could only capture the partial relationship between vegetation and mental health.

Additionally, area-based measures of vegetation, such as the Veg_RF and Tree_OD, are dependent on the spatial resolution of the satellite or aerial image. The data providers noted this limitation for the Tree_OD dataset by mentioning that some features (trees) were missed due to their locations near tall buildings and in deep shadows (Open Data Portal Toronto, 2019). Unfortunately, this problem persists for any vegetation measure that is based on the visual interpretation of aerial images. In highly urbanized settings such as Toronto, with a marked presence of settlements that could reduce the visibility of trees and surrounding vegetation patches, such area-based measures of vegetation might not be suitable for health-based studies. Furthermore, the visual interpretation process is also subjected to the interpretation of the user or the ability to identify different structures of vegetation (tall trees in protected areas, shrubs and bushes in parks, and ornamental plants in gardens and rooftops) in the image. Consequently, this type of dataset might underestimate the vegetation content in the area and the surrounding greenness, as evidenced by the results when visually comparing the raster images of satellite-based vegetation measures (EVI, NDVI, SAVI and Veg_RF) with the area-based measure of tree cover (Tree_OD).

Although the visual interpretation process could be automated through the application of powerful machine learning ensembles such as random forest classifiers to capture vegetation cover, a high degree of landscape heterogeneity, such as that present in an urban setting, could preclude the accurate detection of different types of vegetation in the area (Abdullah et al., 2019; Aplin, 2003). Therefore, land cover classification via RF could be impaired due to the medium to low-resolution of Landsat images (30 m), leading to spectral confusions and problems in differentiating vegetation from other land cover classes (such as a green-colored building or a tennis court) (Abdullah et al., 2019; Aplin, 2003). This misclassification may lead to either over- or under-estimation of the vegetation cover in an area. As a result, the association being detected using such a misclassified dataset would be devoid of the actual relationship between vegetation and mental health disorders. However, the accuracy assessments revealed that the RF model in this study had an accuracy of over 90% for the land cover classification, so the over- and under-estimation should not be a problem for this study. In that case, the inability of Veg_RF to capture the density and biomass conditions of vegetation cover or people's actual exposure to surrounding vegetation could be the actual reason for the differences observed in the results of Bayesian models using Veg_RF and the vegetation indices (EVI and SAVI).

Contrary to the area-based measures, satellite-based vegetation indices such as EVI, NDVI and SAVI can measure both the density and quality or health conditions of the vegetation cover. This is because their values vary based on the chlorophyll content, variations in canopy cover, and canopy architectures (Huete, 1988; Jensen, 2009; Matsushita, Yang, Chen, Onda, & Qiu, 2007). For example, the values of the vegetation indices increase when there are more leaves and more photosynthetic activities in the vegetation patch, which are the measures of density (leaves) and greenness, respectively. Therefore, using these indices can help accurately capture the relationship between the surrounding greenness and poor mental health outcomes (Markevych et al., 2017), as the number of mental health disorders cases is allowed to vary by both the density and health of the surrounding vegetation cover. This could have led to the differences in the results of Bayesian models from the vegetation indices (EVI and SAVI) and area-based measures (Veg_RF and Tree_OD).

Surprisingly, the models for NDVI did not yield any statistically significant association with any of the psychotic or non-psychotic disorders. This could be explained in terms of the computational differences between NDVI and the other two indices. First, NDVI and SAVI are computationally similar, but SAVI could be considered as a modified form of NDVI, where the NDVI is corrected for the influence of soil brightness (USGS, 2019c). The principle of this correction originates from the fact that background brightness from surfaces such as soil may interact with the radiations reflected towards the sensor (satellite) from overlying vegetation canopy (Huete, 1988). Depending upon the canopy and sub-surface scattering, this may result in complex soil surface-vegetation interactions, which affect the values of NDVI. Therefore, in an urbanized area such as Toronto, there could be substantial background noise from the different built-up surfaces such as bitumen covered roads, concrete pavements, brick surfaces and gravel-covered rooftops (Figure 5(b)). The correction for this background noise could have led SAVI to detect the vegetation content in each neighborhood more accurately than NDVI and, thus, better capture the association between surrounding vegetation and mental health disorders.

Second, the natural atmospheric conditions in urban areas are disrupted by pollutants from vehicles and commercial sites (Cleugh & Grimmond, 2011). Additionally, urban morphology such as tall buildings, surface roughness, and low heat capacity of materials such as concrete can affect the wind flow and both vertical and horizontal distributions of these pollutants in the atmosphere (Vallero, 2014). Hence, urban areas like Toronto can be subjected to substantial atmospheric perturbations and resistances from the aerosols and pollutants in the atmosphere. These atmospheric disturbances can affect the transmittance of the red band through the atmosphere to the satellite and so, can influence the NDVI or SAVI values. The EVI can overcome this problem and can adjust for the atmospheric disturbances by using the atmosphere-sensitive blue band to correct the affected red band for atmospheric influences (Huete et al., 2002). EVI is also adjusted for canopy background noise through the canopy signal decoupling process, which makes it very sensitive to vegetation greenness. The decoupling process allows different forms of vegetation to be captured by minimizing the covering effect of large overlying vegetation (Eamus, Huete, & Yu, 2016; Huete et al., 2002; USGS, 2019a). These two factors (atmospheric disturbances and canopy background noise) could have led to the differences in the detection of vegetation content, especially between that of NDVI and EVI in the study area. Although this study could not find any notable differences between the vegetation cover detected by EVI and SAVI, the atmospheric perturbations and the canopy background noise could cause a substantial difference between these two indices in other urban areas, depending on the exact geophysical settings of the studied city. Therefore, it is recommended that future mental health studies empirically check and validate the performances of these two indices in the detection of the surrounding vegetation.

However, considering the urban geophysical settings of Toronto and the potential atmospheric and environmental disturbances, as observed from the LULC map produced using RF and also the Google Earth images in 2016, EVI was preferred over SAVI to model the age and sex-specific effects of vegetation cover and to map the relative risks. The results from the LULC model suggested that during 2016 the study area had an intermediate level of vegetation cover (~ 25% of total area), at which canopy background noise can have substantial effects in the detection of vegetation (Eamus et al., 2016). In this regard, the EVI, as opposed to NDVI or SAVI, was able to adjust this background noise and could be considered as more robust at capturing the relationship between vegetation and the different types of mental health disorders. Therefore, the results from Bayesian models indicate that the vegetation indices, which can incorporate the urban factors affecting the detection of vegetation, could be more suitable for analyzing the relationship of vegetation with mental health in population-based studies. Thus, the findings indicate that it is imperative to consider the type of study area (urban, peri-urban or rural) while selecting the vegetation indices for mental health studies.

The results of this study showed that the vegetation (EVI and SAVI) was negatively associated with psychotic and non-psychotic disorders in Toronto after adjusting for material deprivation, residential instability, dependence, ethnic concentration, substance use disorders and the unmeasured covariates. Comparing the results from the models having the same vegetation measure but different mental health outcomes, the vegetation was found to affect the psychotic disorder cases relatively more than non-psychotic disorders. These findings are quite consistent with studies that studied this relationship and had found that vegetation can positively affect patients with severe mental health disorders, such as affective (mood) and psychotic disorders (Bielinis et al., 2020; Chen et al., 2018). When patients with affective disorders were treated with forest therapies, where the patients had engaged in recreational activities in the nearest suburban forest (with dense vegetation cover), positive effects on 'confusion' and 'depression' were noticed. Similarly, for patients with psychotic disorders, there were significant improvements in the major symptoms of schizophrenia, such as anxiety, dejection, and confusion.

5.2 The association between vegetation and the age and sex-specific mental health disorders

The results further suggest that the mental health disorders are associated with vegetation, for males aged 0-19 and 20-44 years and for females in the age group, 20-44. For the Toronto area, this study could not find any association between vegetation and mental health disorders for older adults from the other age groups (45-64 and 65+ years). These findings are quite consistent with the results obtained from past studies on people from similar age groups but from different study settings. For example, Lee, Kim and Ha (2019) analyzed the association between neighborhood greenness in children's residential areas in South Korea and their neurobehavioral health and found that the higher surrounding greenness was associated with improved neurobehavioral health. In their study, improved mental health conditions in the domains such as reduced aggressive behavior, improved attention, and reduced effects of attention deficit hyperactivity disorder (ADHD) were found for children aged 6 to 18 years with the increase in surrounding greenness (Lee, Kim, & Ha, 2019). The associations were more prominent for the externalizing than the internalizing behaviors and were significant, especially when the greenness was within 1600 m around children's residence. Similarly, Fong et al. (2018), in their review study, had concluded that neighborhood greenness was beneficial for the children's cognitive function and mental health (Fong, Hart, & James, 2018).

However, despite having a similar distribution of mental health disorder cases, unlike males, females in the age group, 0-19, did not show any statistically significant association with EVI. For the models comprising females in this age group, only the ethnic concentration factor was negatively associated with the combined mental health disorder cases. This shows that relative to other variables, there was a marked influence of ethnic concentration in the area that could have affected the distribution of mental health disorder cases for females in this age group. Furthermore, it also can explain the observed non-significant association with EVI. Most conservative immigrant families prefer to keep the girls, especially those who are in their adolescence age, within the confinement of their homes. This was also evidenced in the study conducted by Beiser and Hou (2016), who found that the mental health problems of adolescent females from immigrant and refugee families in Canada are mainly internalizing in nature (Beiser & Hou, 2016). The authors also discussed that mothers who have faced pre-migration adversities and are accustomed to "suffering in the shadows" as a survival strategy may force a similar approach on their children (Beiser & Beiser, 1999; Beiser & Hou, 2016). Similarly, immigrant and refugee families who have migrated from an origin having poor social conditions, where adolescent girls are at significant threat from being sexually harassed or even raped, might be accustomed to keeping girls within the safe vicinity of their homes. These controlling behaviors adopted by the mothers and families of the young girls could lead to fewer interactions with the outside environment compared to boys until they have reached a relatively mature age (for example, 20 years and above). This could have caused the vegetation cover to have significantly less impact on the mental health of girls compared to the boys aged between 0-19 years.

Furthermore, exposure to urban greenery could have a more prominent effect on young adults from the age-group 20-44, compared to older adults, due to the differences in the way people from these age groups develop social cohesion and adopt health-benefiting behaviors. Young people are relatively more physically active, socially engaging and are more likely to adopt beneficial health behaviors than older adults (Johannsen et al., 2008). Consequently, young adults spend more time outdoors than older adults, which exposes them to different levels of urban greenery and vegetation cover. Dzhambov et al. (2018) discussed that people do not only seek greenspaces for physical activities but also to enjoy the restorative capacities of surrounding vegetation cover. Hence, young adults are also likely to be more in contact with the vegetation covers in parks and other green spaces due to using these places as sites for both physical exercise and destressing (Dzhambov, Hartig, Markevych, Tilov, & Dimitrova, 2018a; Dzhambov et al., 2018b). Additionally, past studies suggest that different forms of neighborhood vegetations such as trees can lead to the greater use of outdoor spaces and an increase in social engagements amongst the youth (Coley, Sullivan, & Kuo, 1997; Kuo, Sullivan, Coley, & Brunson, 1998). These social engagements promote social cohesion, which is known to improve mental health conditions in people (Dzhambov et al., 2018a; Markevych et al., 2017).

Although vegetation was negatively associated with the mental health disorder cases in males and females from the age group 20-44, this study could not find any significant association with mental health disorders for people having age 45 years and above (age groups 45-64 and 65+). However, the material deprivation and residential instability covariates showed significant and positive associations with mental health disorders for the age groups 45-64 and 65+. These findings are consistent with the results of Breslin and Mustard (2003), who studied the factors affecting the impact of unemployment on mental health among 6000 young and older adults in Canada. Breslin and Mustard (2003) reported that the respondents aged between 31 to 55 years were substantially affected by unemployment (losing a job) and became psychologically distressed. The distress was so severe in some cases that it often led to clinical depressions (Breslin & Mustard, 2003). In contrast, there was no association between unemployment and mental health conditions amongst young people aged 18-30 years. These findings suggest that compared to young adults, the mental health conditions of older adults could be better explained by socioeconomic variables such as unemployment, which in turn, is closely related to other factors such as income and poverty. Hence, the observed insignificance of the association between vegetation and mental health disorders for the people aged 45-64 and 65+ could be due to the socioeconomic factors substantially dominating the association with mental health disorders and thus, rendering the influence of vegetation as ineffective.

A crucial aspect of using EVI to analyze the age and sex-specific associations between vegetation and mental health disorders could be discussed using the findings from the study conducted by Srugo et al. (2019). Using data from the Ontario Student Drug Use and Health Survey, their study assessed the impact of school-based greenness on mental health conditions among 6,313 students between ages 11-20. However, their findings suggested that there was no association between the quantity of the greenness surrounding the school neighborhood and the student's mental health conditions (Srugo et al., 2019). In this regard, the authors discussed the importance of using vegetation measures that could capture both the quality and the quantity of the surrounding greenery. They also reported that they could not find a statistically significant association between the mental health conditions of the students and the surrounding greenness due to their vegetation measure being unable to capture the quality of the vegetation. A study conducted in Western Australia had similarly concluded that the quality rather than the quantity (or number) of greenspaces was related to the reduced psychological distress of people in their study area (Francis, Wood, Knuiman, & Giles-Corti, 2012). These conclusions support the use of EVI in this study to study the age and sex-specific effects of vegetation on mental health. As discussed earlier, the ability of EVI to capture both the health conditions (the quality) and areal extent of vegetation cover (the quantity) could have led to the differences in results reported by studies such as Srugo et al. (2019) and this research.

5.3 The mental health benefits of the presence of a healthy vegetation cover

If the results are discussed further to explain the role of vegetation in determining the prevalence of mental health disorders, the mental health benefits from vegetation can be broadly categorized into two specific domains, reducing harm and improving restoration capacities (Markevych et al., 2017). Vegetation can help reduce physical harm to the body by improving environmental conditions such as reducing air pollution and exposure to heat and noise. These factors adversely affect the psychological well-being and cognitive development of people, which could later transform into mental health disorders (Dadvand et al., 2016; Dzhambov, 2018; Dzhambov et al., 2018b). The mental health restoration capacity of vegetation could be explained in terms of stress reduction and attention restoration theories (Markevych et al., 2017). The stress reduction theory explains that viewing vegetation and similar natural features can initiate positive thoughts, which, in turn, help control negative thoughts and emotions.

Thus, regular exposure to greenery and the natural environment can help improve stress response and allow people to circumvent negative emotions that deteriorate mental health conditions. In this process, the attention restorative capacities improve as well, as people have better cognition that helps willfully direct attention to the positive aspects of life (Hartig, 2007; Markevych et al., 2017). However, it is also important that future research investigates the extent to which the "green" color in vegetations influences the mental health conditions of the people, more specifically, whether there is any mental health benefit of being exposed to the green color in the vegetation. Based on the existing literature and present knowledge base, it could be concluded that the mental health benefits of vegetation could actually be from the cumulative effects of flora and fauna (rather than the green color) and the aesthetically and psychologically pleasant environment that is created due to the presence of vegetation. In simpler terms, when there is a consistent presence of healthy vegetation cover, birds, animals and other life-forms follow to create a natural and mentally pleasant environment.

5.4 The strengths of this study and recommendations from the findings

The findings of this study provided distinct comparisons between different vegetation measures and showed how their performances might vary in population-level mental health researches. This study emphasized the necessity to select a vegetation measure that can help accurately capture both the quantity and quality of people's exposure to surrounding greenness. In this regard, the results suggest that the satellite-based vegetation indices like EVI, NDVI and SAVI could be particularly useful. Furthermore, this study has also shed light on the importance of incorporating vegetation measures that could account for the atmospheric and environmental disturbances owing to the nature of the study area (for example, urban, peri-urban or rural). Therefore, more sophisticated vegetation indices such as EVI and SAVI could be better choices compared to simpler indices, such as NDVI, for mental health or public health research.

The results from the ecological regressions using BSM provide clear indications that investments on urban vegetation can have tangible health benefit effects, such as improved mental health conditions of the general public. The age and sex-specific analyses revealed that young people could be particularly impacted due to the reduced vegetation cover in an area. Hence, this study created evidence that based on the demographics in an area, investment in vegetation could be extremely helpful in reducing mental health burdens. Furthermore, this research has provided directions that could be extended further to design future studies aiming to understand how long-term investments in urban vegetation could help reduce healthcare costs.

The research incorporated both psychotic and non-psychotic disorders, and so, the findings have captured the gradation of influence exerted by vegetation on the two most common types of mental health disorders. The comparisons of the models using the same vegetation measure but different mental health outcomes such as psychotic and non-psychotic disorders suggest that people suffering from psychotic disorders could be well benefitted from the presence of vegetation in an area. Therefore, this study has also established the need to explore non-conventional and nature-based treatment options, such as ecotherapy, for treating mental health disorders. These treatments could be used as supplements for medical treatments. Future research can explore this by developing longitudinal studies to understand the exact impact of consistent exposure to vegetation cover on the treatment of psychotic and non-psychotic disorder cases.

The analyses conducted in this study quantified the relative contributions of the spatial and non-spatial unmeasured covariates and showed that these latent covariates could significantly explain the prevalence of mental health disorder cases. The models suggested a strong spatial dependence from the unmeasured covariates, which must be addressed during the selection of a modeling technique. The findings also indicated that a spatial modeling approach, incorporating random effect terms to capture the relative contributions of the spatially structured and unstructured unmeasured covariates, could be more realistic and precise compared to non-spatial statistical modeling. Future research can study the association between vegetation and mental health disorders using spatial and non-spatial techniques and can compare the findings to understand how much they differ in epidemiological research.

5.5 Limitations

Despite the strengths, several limitations are present in this study. First, research suggests that the surrounding greenness and exposure to greenness are best captured by the eye-level panoramic imagery of green space (Markevych et al., 2017). However, the process of obtaining such imagery is both time-consuming and expensive. Therefore, this study attempted to demonstrate the performances of the vegetation measures using datasets that are inexpensive and readily available for epidemiological research.

Second, this is an ecological study where the results conform to the findings relevant at the area level for groups of people. The results of the associations need to be interpreted with caution and no individual-level conclusions should be drawn from the findings.

Third, the study has not assessed the performance of an index that utilizes the combined strengths of both EVI and SAVI. Unfortunately, such an attempt is well beyond the scope of this study, as it requires a careful selection of techniques to combine the two indices or perform the adjustments that are conducted during the calculation of these indices.

Fourth, the study did not provide any comparison in terms of the magnitude of the associations with psychotic or non-psychotic disorders for the different vegetation measures. However, this was not possible as the different measures of vegetations were not standardized and so valid comparisons between the magnitude of the associations could not be made. This study did not standardize the various measures of vegetation, which could have allowed this comparison, as the study attempted to retrace the approaches most commonly adopted by public health researchers. The study tried to understand how the selection of any of the five vegetation measures by a random researcher could have affected the detection of a significant association between vegetation and mental health disorders. Therefore, the vegetation measures needed to be used without further modifications and in their original forms, just as they would be commonly used in a public health study.

Regardless of these limitations, this study has taken up the challenge to identify the methodological constraints owing to the selection of different vegetation measures in population-based mental health studies. This research attempted to understand the complex relationship between vegetation and mental health disorders by developing hierarchical models that adjust for potential confounders and unmeasured covariates.

Chapter 6 Conclusions

The increase in global urbanization and the subsequent loss of vegetation covered areas are likely to put millions of people at risk from poor mental health conditions. Unfortunately, due to the disagreements from carefully designed studies, it is still unclear whether reduced vegetation is a significant risk factor for mental health disorders. However, there is a paucity of studies that have assessed the performances of different types of vegetation measures in studying the association between vegetation and mental health disorders. Therefore, through the application of remote sensing, geographic information system and machine learning techniques, three satellite-based indices and two area-based measures of vegetation were used to analyze the relationship between vegetation and psychotic and non-psychotic disorders, after adjusting for material deprivation, ethnic concentration, residential instability, dependence, the rate of substance use disorders and unmeasured (latent) covariates. The results from this analysis were further investigated to select a suitable vegetation measure, which was later employed to study the age and sex-specific effects of the vegetation on mental health disorders. The associations were studied using Poisson-lognormal models under a Bayesian framework. The vegetation was found to be negatively associated with both psychotic and non-psychotic disorders. Results suggest that the satellite-based indices could be better than area-based measures at capturing a significant association with mental health. The findings also indicate that the indices, such as enhanced vegetation index and soil adjusted vegetation index, which are adjusted for atmospheric disturbances, canopy background and soil-brightness, could be particularly useful. The age and sex-specific analyses suggest that the mental health conditions of children and younger adults could be the most adversely affected due to reduced vegetation cover. Additionally, the mapping of the relative risks provided evidence of both macro and micro-level variations in risk from mental health disorders, which could be the focus of targeted public health interventions. The findings from this study are expected to provide critical guidelines on the selection of an appropriate vegetation measure for future population-based mental health studies. The findings could also be helpful for other health research that use such measures to understand the exposure of the general public to surrounding vegetation cover.

References

Abdullah, A. Y. M., Masrur, A., Adnan, M. S. G., Baky, M., Al, A., Hassan, Q. K., & Dewan, A. (2019). Spatio-Temporal Patterns of Land Use/Land Cover Change in the Heterogeneous Coastal Region of Bangladesh between 1990 and 2017. *Remote Sensing, 11*(7), 790.

Annerstedt, M., Jönsson, P., Wallergård, M., Johansson, G., Karlson, B., Grahn, P., . . . Währborg, P. (2013). Inducing physiological stress recovery with sounds of nature in a virtual reality forest—Results from a pilot study. *Physiology & behavior, 118*, 240-250.

Anselin, L. (1990). Spatial dependence and spatial structural instability in applied regression analysis. *Journal of Regional Science, 30*(2), 185-207.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis, 27*(2), 93-115.

Aplin, P. (2003). Comparison of simulated IKONOS and SPOT HRV imagery for classifying urban areas. *Remotely sensed cities*, 23-45.

Arnold, N., Thomas, A., Waller, L., & Conlon, E. (1999). *Bayesian models for spatially correlated disease and exposure data.* Paper presented at the Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting.

Astell-Burt, T., Mitchell, R., & Hartig, T. (2014). The association between green space and mental health varies across the lifecourse. A longitudinal study. *J Epidemiol Community Health, 68*(6), 578-583.

Bangasser, D. A., & Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: neurobiological perspectives. *Frontiers in neuroendocrinology, 35*(3), 303-319.

Beiser, M., & Beiser, M. (1999). *Strangers at the gate: The" boat people's" first ten years in Canada*: University of Toronto Press.

Beiser, M., & Hou, F. (2016). Mental health effects of premigration trauma and postmigration discrimination on refugee youth in Canada. *The Journal of nervous and mental disease, 204*(6), 464-470.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4): Springer.

Bezold, C. P., Banay, R. F., Coull, B. A., Hart, J. E., James, P., Kubzansky, L. D., . . . Laden, F. (2018). The relationship between surrounding greenness in childhood and adolescence and depressive symptoms in adolescence and early adulthood. *Annals of epidemiology, 28*(4), 213-219.

Bielinis, E., Jaroszewska, A., Łukowski, A., & Takayama, N. (2020). The Effects of a Forest Therapy Programme on Mental Hospital Patients with Affective and Psychotic Disorders. *International journal of environmental research and public health, 17*(1), 118.

Bjarnason, T., & Sigurdardottir, T. J. (2003). Psychological distress during unemployment and beyond: social support and material deprivation among youth in six northern European countries. *Social Science & Medicine, 56*(5), 973-985.

Blair, A., Gariépy, G., & Schmitz, N. (2015). The longitudinal effects of neighbourhood social and material deprivation change on psychological distress in urban, community-dwelling Canadian adults. *public health, 129*(7), 932-940.

Borrell, C., Muntaner, C., Solè, J., Artazcoz, L., Puigpinos, R., Benach, J., & Noh, S. (2008). Immigration and self-reported health status by social class and gender: the importance of material deprivation, work organisation and household labour. *Journal of Epidemiology & Community Health, 62*(5), e7-e7.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Breslin, F. C., & Mustard, C. (2003). Factors influencing the impact of unemployment on mental health among young and older adults in a longitudinal, population-based survey. *Scandinavian journal of work, environment & health*, 5-14.

Brown, D. K., Barton, J. L., & Gladwell, V. F. (2013). Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress. *Environmental science & technology, 47*(11), 5562-5569.

Chen, H.-T., Yu, C.-P., & Lee, H.-Y. (2018). The effects of forest bathing on stress recovery: evidence from middle-aged females of taiwan. *Forests, 9*(7), 403.

City of Toronto. (2013). *Sustaining and expanding the urban forest: Toronto's strategic forest management plan*. City of Toronto, Parks, Forestry and Recreation Division …

Cleugh, H., & Grimmond, S. (2011). Urban climates and global climate change. *The Future of the World's Climate (Second Edition)*, 47-76.

Coley, R. L., Sullivan, W. C., & Kuo, F. E. (1997). Where does community grow? The social context created by nature in urban public housing. *Environment and behavior, 29*(4), 468-494.

Collinearity Diagnostics, Model Fit & Variable Contribution. Retrieved from https://cran.r-project.org/web/packages/olsrr/vignettes/regression_diagnostics.html

Currie, S. R., Patten, S. B., Williams, J. V., Wang, J., Beck, C. A., El-Guebaly, N., & Maxwell, C. (2005). Comorbidity of major depression with substance use disorders. *The Canadian Journal of Psychiatry, 50*(10), 660-666.

Cutler, L. B. A. (2004). *Random Forests*. Retrieved from https://www.stat.berkeley.edu/~breiman/RandomForests/

Dadvand, P., Bartoll, X., Basagaña, X., Dalmau-Bueno, A., Martinez, D., Ambros, A., . . . Borrell, C. (2016). Green spaces and general health: roles of mental health status, social support, and physical activity. *Environment international, 91*, 161-167.

Dzhambov, A., Hartig, T., Markevych, I., Tilov, B., & Dimitrova, D. (2018a). Urban residential greenspace and mental health in youth: Different approaches to testing multiple pathways yield different conclusions. *Environmental Research, 160*, 47-59.

Dzhambov, A. M. (2018). Residential green and blue space associated with better mental health: a pilot follow-up study in university students. *Archives of Industrial Hygiene and Toxicology, 69*(4), 340-349.

Dzhambov, A. M., Markevych, I., Hartig, T., Tilov, B., Arabadzhiev, Z., Stoyanov, D., . . . Dimitrova, D. D. (2018b). Multiple pathways link urban green-and bluespace to mental health in young adults. *Environmental Research, 166*, 223-233.

Eamus, D., Huete, A., & Yu, Q. (2016). *Vegetation dynamics*: Cambridge University Press.

Feehan, M., McGee, R., & Williams, S. M. (1993). Mental health disorders from age 15 to age 18 years. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(6), 1118-1126.

Fong, K. C., Hart, J. E., & James, P. (2018). A review of epidemiologic studies on greenness and health: updated literature through 2017. *Current environmental health reports, 5*(1), 77-87.

Francis, J., Wood, L. J., Knuiman, M., & Giles-Corti, B. (2012). Quality or quantity? Exploring the relationship between Public Open Space attributes and mental health in Perth, Western Australia. *Social Science & Medicine, 74*(10), 1570-1577.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern recognition letters, 27*(4), 294-300.

Glazier, R. H., Gozdyra, P., Kim, M., Bai, L., Kopp, A., Schultz, S. E., & Tynan, A.-M. (2018). Geographic Variation in Primary Care Need, Service Use and Providers in Ontario, 2015/16. Toronto, ON: Institute for Clinical Evaluative Sciences.

Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological applications, 10*(6), 1861-1870.

Han, B., Gfroerer, J. C., Colliver, J. D., & Penne, M. A. (2009). Substance use disorder among older adults in the United States in 2020. *Addiction, 104*(1), 88-96.

Hartig, T. (2007). Three steps to understanding restorative environments as health resources. In *Open space: People space* (pp. 183-200): Taylor & Francis.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of Environment, 83*(1-2), 195-213.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of Environment, 25*(3), 295-309.

Huynh, Q., Craig, W., Janssen, I., & Pickett, W. (2013). Exposure to public natural space as a protective factor for emotional well-being among young people in Canada. *BMC public health, 13*(1), 407.

James, P., Banay, R. F., Hart, J. E., & Laden, F. (2015). A review of the health benefits of greenness. *Current epidemiology reports, 2*(2), 131-142.

Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective 2/e*: Pearson Education India.

Jiang, B., Deal, B., Pan, H., Larsen, L., Hsieh, C.-H., Chang, C.-Y., & Sullivan, W. C. (2017). Remotely-sensed imagery vs. eye-level photography: Evaluating associations among measurements of tree cover density. *Landscape and urban planning, 157*, 270-281.

Johannsen, D. L., DeLany, J. P., Frisard, M. I., Welsch, M. A., Rowley, C. K., Fang, X., . . . Ravussin, E. (2008). Physical activity in aging: comparison among young, aged, and nonagenarian individuals. *Journal of Applied Physiology, 105*(2), 495-501.

Jones, P. B. (2013). Adult mental health disorders and their age at onset. *The British Journal of Psychiatry, 202*(s54), s5-s10.

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*(9), 1548-1555.

Khan, A. M., Urquia, M., Kornas, K., Henry, D., Cheng, S. Y., Bornbaum, C., & Rosella, L. C. (2017). Socioeconomic gradients in all-cause, premature and avoidable mortality among immigrants and long-term residents using linked death records in Ontario, Canada. *J Epidemiol Community Health, 71*(7), 625-632.

Kuo, F. E., Sullivan, W. C., Coley, R. L., & Brunson, L. (1998). Fertile ground for community: Inner-city neighborhood common spaces. *American Journal of Community Psychology, 26*(6), 823-851.

Law, J., & Haining, R. (2004). A Bayesian approach to modeling binary data: The case of high-intensity crime areas. *Geographical analysis, 36*(3), 197-216.

Law, J., Haining, R., Maheswaran, R., & Pearson, T. (2006). Analyzing the relationship between smoking and coronary heart disease at the small area level: a Bayesian approach to spatial modeling. *Geographical analysis, 38*(2), 140-159.

Law, J., & Perlman, C. (2018). Exploring geographic variation of mental health risk and service utilization of doctors and hospitals in Toronto: A shared component spatial modeling approach. *International journal of environmental research and public health, 15*(4), 593.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*: Chapman and Hall/CRC.

Lee, M., Kim, S., & Ha, M. (2019). Community greenness and neurobehavioral health in children and adolescents. *Science of the Total Environment, 672*, 381-388.

LeSage, J. P. (1997). Regression analysis of spatial data. *Journal of Regional Analysis and Policy, 27*(1100-2016-89650), 83-94.

Ma, J., Li, C., Kwan, M.-P., & Chai, Y. (2018). A multilevel analysis of perceived noise pollution, geographic contexts and mental health in Beijing. *International journal of environmental research and public health, 15*(7), 1479.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician, 36*(3a), 158-160.

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A. M., . . . Nieuwenhuijsen, M. J. (2017). Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environmental Research, 158*, 301-317.

Martellozzo, F., Ramankutty, N., Hall, R. J., Price, D. T., Purdy, B., & Friedl, M. A. (2015). Urbanization and the loss of prime farmland: A case study in the Calgary–Edmonton corridor of Alberta. *Regional environmental change, 15*(5), 881-893.

Matheson FI, & van Ingen T. (2018). *2016 Ontario marginalization index: user guide. Toronto, ON: St. Michael's Hospital; 2018. Joint publication with Public Health Ontario.*

Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: a case study in high-density cypress forest. *Sensors, 7*(11), 2636-2651.

Mckenzie, S. K., Gunasekara, F. I., Richardson, K., & Carter, K. (2014). Do changes in socioeconomic factors lead to changes in mental health? Findings from three waves of a population based panel study. *J Epidemiol Community Health, 68*(3), 253-260.

Melis, G., Gelormino, E., Marra, G., Ferracin, E., & Costa, G. (2015). The effects of the urban built environment on mental health: A cohort study in a large northern Italian city. *International journal of environmental research and public health, 12*(11), 14898-14915.

Mitchell, R., & Popham, F. (2008). Effect of exposure to natural environment on health inequalities: an observational population study. *The Lancet, 372*(9650), 1655-1660.

Morgan, V. A., Castle, D. J., & Jablensky, A. V. (2008). Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) Disorders. *Australian and New Zealand Journal of Psychiatry, 42*(1), 74-82.

Ontario Community Health Profiles Partnership. (2019). *Data — LHIN 7 (Toronto Central and City of Toronto) Neighbourhoods, Ontario Sub-Regions and LHINs*. Retrieved from http://www.ontariohealthprofiles.ca/dataTablesON.php?varTab=HPDtbl&select1=7

Open Data Portal Toronto. (2019). *About Topographic Mapping – Treed Area*. Retrieved from https://open.toronto.ca/dataset/topographic-mapping-treed-area/

Peñuelas, J., & Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in plant science, 3*(4), 151-156.

Pieterse, A. L., Todd, N. R., Neville, H. A., & Carter, R. T. (2012). Perceived racism and mental health among Black American adults: A meta-analytic review. *Journal of Counseling Psychology, 59*(1), 1.

Puissant, A., Rougier, S., & Stumpf, A. (2014). Object-oriented mapping of urban trees using Random Forest classifiers. *International Journal of Applied Earth Observation and Geoinformation, 26*, 235-245.

Reid, M. C., & Anderson, P. A. (1997). Geriatric substance use disorders. *Medical Clinics of North America, 81*(4), 999-1016.

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Social Science & Medicine, 90*, 24-31.

Robertson, C., Nelson, T. A., MacNab, Y. C., & Lawson, A. B. (2010). Review of methods for space–time disease surveillance. *Spatial and spatio-temporal epidemiology, 1*(2-3), 105-116.

Rugel, E. J., Carpiano, R. M., Henderson, S. B., & Brauer, M. (2019). Exposure to natural space, sense of community belonging, and adverse mental health outcomes across an urban region. *Environmental Research, 171*, 365-377.

Rugel, E. J., Henderson, S. B., Carpiano, R. M., & Brauer, M. (2017). Beyond the Normalized Difference Vegetation Index (NDVI): developing a natural space index for population-level health research. *Environmental Research, 159*, 474-483.

Saraceno, B., Levav, I., & Kohn, R. (2005). The public mental health significance of research on socio-economic factors in schizophrenia and major depression. *World psychiatry, 4*(3), 181.

Sasaki, A., Vega, W. C. d., & McGowan, P. O. (2013). Biological embedding in mental health: an epigenomic perspective. *Biochemistry and Cell Biology, 91*(1), 14-21.

Satcher, D. (2001). Mental health: Culture, race, and ethnicity—A supplement to mental health: A report of the surgeon general. In: US Department of Health and Human Services.

Sharpe, D. M., Stearns, F., Leitner, L. A., & Dorney, J. R. (1986). Fate of natural vegetation during urban development of rural landscapes in southeastern Wisconsin. *Urban Ecology, 9*(3-4), 267-287.

Sheppard, A. J., Salmon, C., Balasubramaniam, P., Parsons, J., Singh, G., Jabbar, A., . . . Dunn, J. (2012). Are residents of downtown Toronto influenced by their urban neighbourhoods? Using concept mapping to examine neighbourhood characteristics and their perceived impact on self-rated mental well-being. *International journal of health geographics, 11*(1), 31.

Simoni-Wastila, L., & Yang, H. K. (2006). Psychoactive drug abuse in older adults. *The American journal of geriatric pharmacotherapy, 4*(4), 380-394.

Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote sensing of Environment, 75*(2), 230-244.

Srugo, S. A., de Groh, M., Jiang, Y., Morrison, H. I., Hamilton, H. A., & Villeneuve, P. J. (2019). Assessing the impact of school-based greenness on mental health among adolescent students in Ontario, Canada. *International journal of environmental research and public health, 16*(22), 4364.

Takayama, N., Saito, H., Fujiwara, A., & Horiuchi, M. (2017). The effect of slight thinning of managed coniferous forest on landscape appreciation and psychological restoration. *Progress in Earth and Planetary Science, 4*(1), 17.

Tomita, A., Vandormael, A. M., Cuadros, D., Di Minin, E., Heikinheimo, V., Tanser, F., . . . Burns, J. K. (2017). Green environment and incident depression in South Africa: a geospatial analysis and mental health implications in a resource-limited setting. *The Lancet Planetary Health, 1*(4), e152-e162.

Toronto Community Health Profiles. *Toronto Health Profiles Information about TCHPP Geographies—Definitions, Notes and Historical Context*. Retrieved from http://www.torontohealthprofiles.ca/a_documents/aboutTheData/0_2_Information_About_TCHPP_Geographies.pdf

United Nations, D. o. E. S. A. (2018). *68% of the World Population Projected to Live in Urban Areas by 2050, Says UN*.

USGS-EarthExplorer. *USGS- EarthExplorer* Retrieved from https://earthexplorer.usgs.gov/

USGS. (2017). *Landsat 8 Data Users Handbook - Section 5*. Retrieved from https://landsat.usgs.gov/landsat-8-l8-data-users-handbook-section-5

USGS. (2019a). *Landsat Surface Reflectance-Derived Spectral Indices - Landsat Enhanced Vegetation Index*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-enhanced-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con

USGS. (2019b). *Landsat Surface Reflectance-Derived Spectral Indices - Landsat Normalized Difference Vegetation Index*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-normalized-difference-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con

USGS. (2019c). *Landsat Surface Reflectance-Derived Spectral Indices - Landsat Soil Adjusted Vegetation Index*. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-soil-adjusted-vegetation-index

Vallero, D. A. (2014). *Fundamentals of air pollution*: Academic press.

Villeneuve, P., Ysseldyk, R., Root, A., Ambrose, S., DiMuzio, J., Kumar, N., . . . Li, X. (2018). Comparing the normalized difference vegetation index with the Google street view measure of vegetation to assess associations between greenness, walkability, recreational physical activity, and health in Ottawa, Canada. *International journal of environmental research and public health, 15*(8), 1719.

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors, 2017*.

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture, 13*(6), 693-712.

Appendices

Appendix A: Details of the mental health disorder dataset

Enrollment, Access, Continuity and Mental Health Gaps in Care

Dataset: Prevalence of mental health disorders and substance use by age, sex, and enrolled/non-enrolled status in City of Toronto and LHIN 7 by neighbourhood, 2015/16

ICES Project No.: 2018 0900 992 000

Data sources: A number of data sources, all held at ICES, were used to prepare this dataset. The sources and the type of data extracted are listed below:

a) ***OHIP - Ontario Health Insurance Plan:*** Health care provider claims
b) ***RPDB - Registered Persons Database:*** Ontario population and OHIP eligibility data
c) ***CPDB - Corporate Provider Database:*** Physician and group data from the Ministry of Health
d) ***IPDB - ICES Physician Database:*** Annual physician demographics, specialization and workload
e) ***CONTACT:*** Yearly health services contact and RPDB eligibility summaries
f) ***CAPE - Client Agency Program Enrollment:*** Registry of patients enrolled in primary care groups
g) ***CIC - Immigration, Refugees and Citizenship Canada (IRCC)'s Permanent Resident Database:*** Ontario portion of IRCC's Permanent Resident Database, including immigration application records for people who initially applied to land in Ontario

Study period: Fiscal 2015 (April 1, 2015 to March 31, 2016)

Study population: All Ontario permanent residents who are eligible for coverage under the publically-funded Ontario Health Insurance Plan (OHIP) on March 31, 2016

Inclusion / Exclusion Criteria:

a) **Inclusion criteria:**
All Ontario permanent residents eligible for OHIP coverage on March 31, 2016.
b) **Exclusion criteria:**
1. Invalid IKN
2. Death before March 31st, 2016
3. No contact within 8 years prior to March 31, 2016
4. Age>105 years
5. People living in long-term care and complex continuing care during the study period

Indicators: Mental health disorders were measured using outpatient visit/claim (OHIP)

Numerator: The number of individuals who had OHIP claims for the mental health conditions listed in Section 3.2.1, Table 1.
Denominator: Total number of people who had a valid health card number and were alive on March 31, 2016

Appendix B: Testing for the spatial autocorrelation in mental health data using global Moran's I analysis

The global Moran's I test was executed on the age and sex-standardized rates (per 1000 population) of both sexes for psychotic and non-psychotic disorders and on the crude rates (per 1000 population) for individual age-group data of the combined mental health disorders. This analysis helped to understand whether there is a statistically significant spatial autocorrelation in the data. Based on the results of this test, the modeling technique for studying the association was selected.

The test was repeated for each of the psychotic, non-psychotic and the combined mental health disorder variables and the first-order Queen's case contiguity was used to define the spatial weight matrix. This weight matrix helped to identify the adjacent neighbors of each neighborhood in the Toronto area and evaluated the similarity and dissimilarity between the values of each neighborhood and its corresponding neighbors. The global Moran's I values range from -1 to + 1, where a highly negative value (Moran's I → -1) will correspond to a perfect dispersion of the mental health disorder cases and a value of 0 will correspond to a random distribution. In contrast, a highly positive value (Moran's I → +1) will indicate a marked spatial autocorrelation in the data and that the like values (high or low) are highly clustered together. The pseudo-p-values, which assessed the significance of the Moran's I values, were generated using 999 permutations. The results of the tests are summarized in Appendix Table 1.

Appendix Table 1: Results of the test for detecting spatial autocorrelation in the data (global Moran's I test)

Type	Moran's I value	p-value	Pattern
Psychotic disorder	0.508	0.001	Moderately clustered
Non-psychotic disorder	0.770	0.001	Highly clustered
Combined mental health disorder			
0-19	0.491	0.001	Moderately clustered
20-44	0.702	0.001	Highly clustered
45-64	0.717	0.001	Highly clustered
65+	0.696	0.001	Highly clustered

Appendix Table 1 shows that two out of the six variables in the retrieved datasets show moderate clustering, while four out of the six variables show the presence of high clustering. A Moran's I value close to 0.5 and 0.7 was considered moderately and highly clustered, respectively. These results confirmed the need to use a spatial modeling technique that adjusts for spatial autocorrelation in the data.

Appendix C: Detailed list of the indicators used to create the four major dimensions in the Ontario Marginalization Index

Appendix Table 2: The four major dimensions of OMI with their indicators

Material Deprivation	Ethnic Concentration	Residential Instability	Dependency
Proportion of the population aged 20+ without a high-school diploma	Proportion of the population who are recent immigrants (arrived in the past 5 years)	Proportion of the population living alone	Proportion of the population who are aged 65 and older
Proportion of families who are lone parent families	Proportion of the population who self-identify as a visible minority	Proportion of the population who are not youth (age 5-15)	Dependency ratio (total population 0-14 and 65+ / total population 15 to 64)
Proportion of total income from government transfer payments for population aged 15+		Average number of persons per dwelling	Proportion of the population not participating in labour force (aged 15+)
Proportion of the population aged 15+ who are unemployed		Proportion of dwellings that are apartment buildings	
Proportion of the population considered low-income		Proportion of the population who are single/divorced/widowed	
Proportion of households living in dwellings that are in need of major repair		Proportion of dwellings that are not owned	
		Proportion of the population who moved during the past 5 years	

Appendix D: Pearson correlation coefficient and multicollinearity tests of the Ontario Marginalization Index (OMI) variables

Prior to running the Pearson correlation coefficient and multicollinearity tests, the individual relationships amongst the four OMI dimensions were assessed using graphical representations. Appendix Figure 1 shows the inter-relationships amongst the OMI variables. For the most part, the graphs indicate that there is no notable linear association between the variables.

A) Pearson correlation coefficient test

The Pearson correlation coefficient test was used to assess the linear association between two or more OMI variables. Highly positive values correspond to positive associations between the tested variables and highly negative values correspond to negative associations. The correlation coefficient values range from -1 to +1. In general, the correlation coefficients having absolute values:

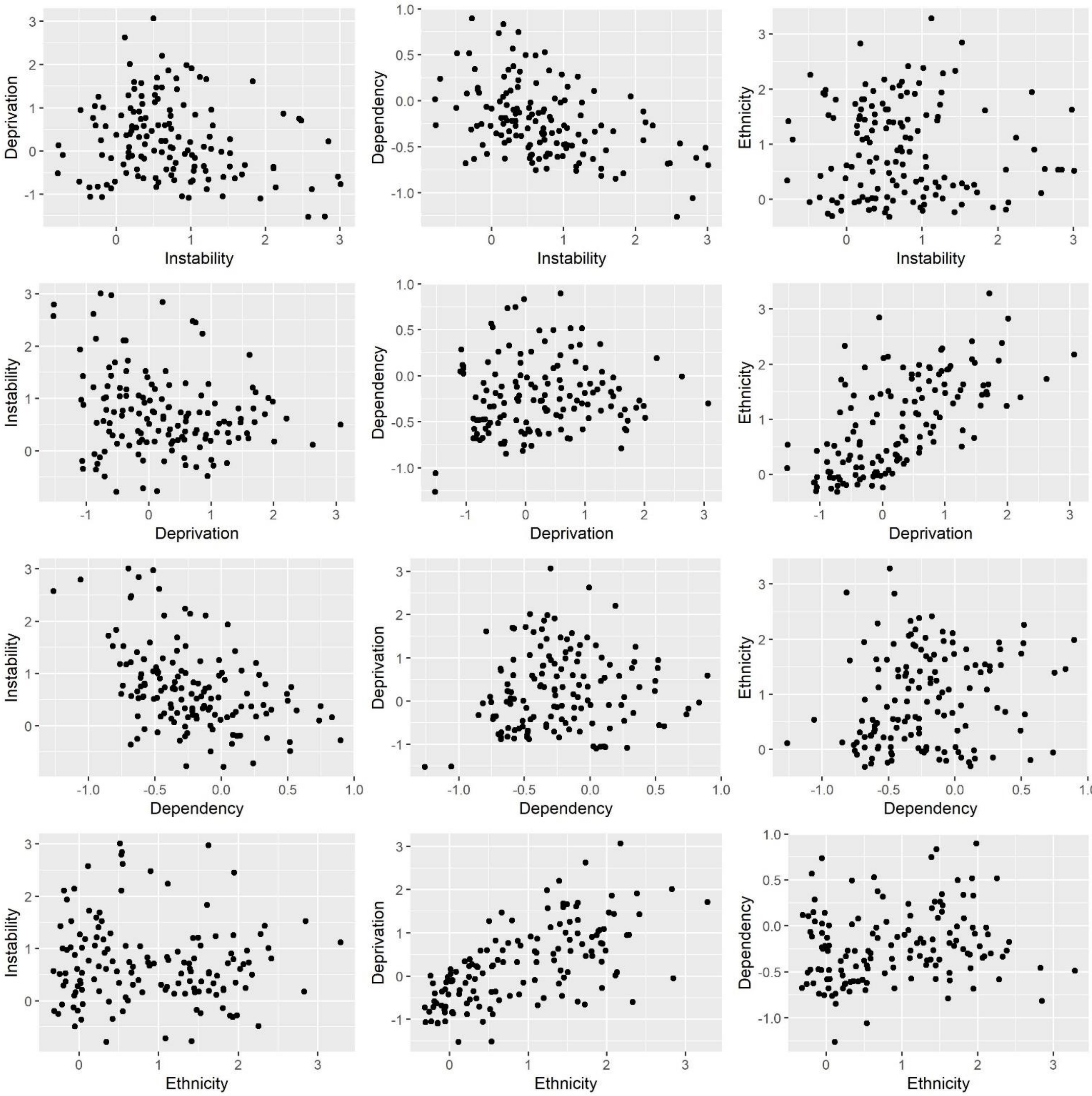
a) 0 to 0.25 represent a low correlation
b) 0.25 to 0.50 represent a moderately low correlation
c) 0.50 to 0.75 represent a moderate correlation
d) 0.75 to 1 represent a high correlation

The results tabulated in Appendix Table 3 indicate that except for correlations between instability and dependency, and between deprivation and ethnic concentration, the magnitudes of the remaining correlations were very small. The correlations coefficient value for instability and dependency was moderately low (> 0.5). The correlation coefficient for deprivation and ethnic concentration indicates a moderate correlation (> 0.75).

Appendix Table 3: The result of the Pearson correlation coefficient test on the four OMI variables

Variables	**Instability**	**Deprivation**	**Dependency**	**Ethnic concentration**
Instability	**1***	**-0.200****	**-0.458***	-0.056
Deprivation		**1***	0.101	**0.649***
Dependency			**1***	**0.175****
Ethnic concentration				**1**

** significant at $p < 0.01$*
***significant at $p < 0.05$*



Appendix Figure 1: Showing the interrelationships amongst the four OMI variables

B) Multicollinearity test:

As the Pearson correlation coefficient test could only evaluate the linear relationship between two variables at a single time, a multicollinearity test was conducted using the *olsrr* package in R (https://cran.r-project.org/web/packages/olsrr/olsrr.pdf). Multicollinearity occurs when the predictor or independent variables in a regression model are strongly linearly correlated with each other or show high inter-associations (Mansfield & Helms, 1982).

For this study, the Tolerance and the Variance Inflation Factors (VIF) were assessed to understand the extent of multicollinearity amongst the variables.

The details of the test statistics and their results are provided below:

1) **Tolerance (1- R^2):**
The tolerance is calculated by regressing one of the four OMI variables (the k^{th} predictor) on rest of the three OMI variables. The R^2 value (R_k^2) is computed and then subtracted from 1 to give the tolerance.

The tolerance indicates the proportion of variance in the dependent OMI variable (Y) that cannot be explained by the remaining three independent OMI variables (X_1, X_2 and X_3). The computational process involves the use of the regression models, detailed in Appendix Table 4. The results of the multicollinearity test are tabulated in Appendix Table 5.

For example, the tolerance value for deprivation indicates that about 50% of the variance in deprivation cannot be explained by the remaining three variables (instability, dependency and ethnic concentration). Similarly, the deprivation, dependency and ethnic concentration variables cannot explain 70% of the variance in the instability dimension of OMI.

The high tolerance values indicate that the four OMI variables are relatively unique and cannot be linearly predicted from one another with sufficient details.

2) **Variance Inflation Factors (VIF):**
The VIF equals to $\frac{1}{\text{Tolerance}}$ and evaluates the inflation in the variances of the parameter estimates due to the collinearities amongst the predictor or independent variables.

In general, variables with VIF values greater than 4 require further investigation to understand their relative contributions in the model and VIF values greater than 10 must have to be corrected using statistical techniques such as the Principal Component Analysis (*Collinearity Diagnostics, Model Fit & Variable Contribution*).

The VIF values for each of the tested variables are well below 4, indicating that there are insufficient collinearities amongst the remaining three predictor variables to inflate the parameter estimates. This confirms the conclusion from the evaluation of tolerance values that the variables are relatively unique to each other and, when added together in a regression model, should not demonstrate sufficient multicollinearity to bias the regression results.

As the results of the Pearson correlation coefficient and multicollinearity tests did not yield signs of notable correlation and multicollinearity amongst the variables, more sophisticated tests of multicollinearity, such as the evaluation of the eigenvalues to assess the relative contributions of the OMI variables in a regression model, were not necessary.

Appendix Table 4: The models used to generate the multicollinearity test statistics

	Y	**X_1**	**X_2**	**X_3**
Model 1:	Instability	Deprivation	Dependency	Ethnic concentration
Model 2:	Deprivation	Instability	Dependency	Ethnic concentration
Model 3:	Dependency	Instability	Deprivation	Ethnic concentration
Model 4:	Ethnic concentration	Instability	Deprivation	Dependency

Appendix Table 5: Results of the multicollinearity test

Model	**Variable tested**	**R^2**	**Tolerance**	**VIF**
1	Instability	0.260	0.740	1.352
2	Deprivation	0.458	0.542	1.844
3	Dependency	0.246	0.754	1.326
4	Ethnic concentration	0.453	0.547	1.827

Appendix E: Flowchart showing the overall methodology of the research