Substance Abuse and Health: A Structural Equation Modeling Approach to Assess Latent Health Effects

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Mathematics and Statistics

University of Saskatchewan

Saskatoon

By

Naorin Islam

Permission to Use

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

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

Head of the Department of Mathematics and Statistics
Room 142 McLean Hall
106 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5E6

Abstract

Background of the Study. Repeated use of a substance (alcohol or drug) may lead to mental and physical sickness, personality changes, insomnia, nausea, mood swings and other disturbances. The number of people addicted to alcohol and/or drug has been increasing every year at an alarming rate. Although the extent of abuse is not directly measurable (i.e., latent), statistical techniques allow us to describe such a hypothetical construct using available information.

Objective. There are many factors potentially associated with substance abuse (e.g., smoking, education, cultural background). Although these variables are readily available in many studies, the cause (e.g., a measure of drug or alcohol abuse) is latent, with the observed variables being its manifestations. A measure of a latent health factor index could also be of particular interest. In this study, we investigate the effects of socio-demographic variables on substance (drug and alcohol) abuse and health in the Canadian population. In particular, the objective is to address the following questions: (a) What would be a reasonable hypothesis to explain causes of substance abusive behavior (i.e., cause and effect relationship)? (b) What model would adequately describe the cause-and-effect relationship between the observed variables and health and substance-related latent variables? (c) What covariates are significantly associated with alcohol and drug abusive environments and health status?

Method. To describe the cause-and-effect relationship among substance abuse, health and socio-demographic variables, we consider structural equation modeling. One of the appealing features of this technique is that it provides a concise assessment of complex relationships. The idea is to formulate a hypothesis regarding such relationships based on prior knowledge about the problem at hand, and then evaluate this hypothesis using statistical techniques. The main goal is to develop a model/hypothesis which can adequately describe the interrelationships among these variables.

Summary Results. The study is based on a survey conducted by Health Canada. We consider 2012 survey data for Saskatchewan and Manitoba, and then develop models to describe the complex relationships among three hypothetical constructs (drug and alcohol abusive environments and heath) and socio-demographic variables. One of the important findings of the study is that an increase in the severity of drug abusive environment may worsen the health of individuals. Another interesting finding is that smoking has no direct effect on health, but it may lead to an environment (alcohol or drug abusive) that could have negative impact on health. Based on our findings, we conclude that substance abuse may significantly deteriorate health. This research will provide policy-makers as well as the public with an understanding of the extent of impacts of substance abuse and relevant socio-demographic variables on health.

Acknowledgements

I am humbly grateful to the Almighty who blessed me the opportunity for higher studies. I cannot express enough thanks to my supervisors, Dr. Shahedul Khan and Dr. Longhai Li. I would like to express my deepest gratitude and appreciation for their patient supervision, who consistently allowed this paper to be my own work, but steered me in the right direction whenever I needed it. Without their precious support, it would not have been possible to complete my degree. I could not have imagined having better mentors and advisors for my Masters study.

I would also like to express my gratitude to Dr. Raj Srinivasan and Dr. Juxin Liu on my advisory committee for their encouragement and insightful comments.

I am grateful to all the professors, graduate students and stuff in the Department of Mathematics & Statistics specially Graduate Program Administrator Kyla Denton, Dr. Lawrence Chang and graduate student Yue Dong and April Liu. I am thankful to the Department of Mathematics & Statistics for the funding provided throughout my Masters program.

Finally, I sincerely thank my parents, friends and specially my husband for their unconditional love, support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Although an uncontrolled use of drugs and/or an increased consumption of alcohol may pose a significant risk to human health, people are sometimes inclined towards alcohol and illicit drugs without realizing the possible consequences (CCSA 2007). Psychological, behavioral and biological changes in adolescence sometimes lead to the first use of alcohol and drug in the young population (NIDA 2007). The use of these substances at an early stage of life lead to addiction, which in turn may complicate health conditions in adolescence or adult years (Bootzin and Stevens 2005). Recent studies suggest that unawareness of the risks associated with drug and alcohol use plays a major role in addiction (CCSA 2007). There are many other risk factors associated with substance (drug and alcohol) abuse, including age, sex, smoking, education and culture. Although information on these variables is readily available in many studies, the response variables (e.g., measures of drug and alcohol abuse) are often not directly measurable but are latent, with the observed variables being their manifestations. A measure of overall health status could also be of particular interest. This variable is possibly manifested by latent variables like drug and alcohol abusive environment, and therefore is also latent. In this study, we investigate the effects of socio-demographic variables on substance abuse and health status in the Canadian population. We consider the structural equation modeling for this purpose, which is a powerful statistical tool for causal inference among the observed and latent variables. This would also help us to understand the interrelationships among the variables under study, and to develop hypotheses based on statistical methods for prediction.

1.1 Background of the Study

Substance abuse generally starts with drinking alcohol, which often leads to smoking and subsequent use of cannabis, marijuana and other drugs (Kandel and Faust 1975). That is, people who use cocaine, marijuana, and cannabis usually also consume alcohol. According to Leyton and Stewart (2014), alcohol is the most widely used legal substance, whereas cannabis is the most widely used illegal drug in Canadian youths. Pharmaceutical drug abuse is also common, though the majority of people abusing pharmaceutical drugs are unaware of the possible consequences (Boyd et al. 2005). Note that some socio-demographic variables are also associated with alcohol and drug-related issues, including household income, wealth, cultural background and education (Patrick et al. 2012).

1.1.1 Global Substance Abuse

Alcohol and drug abuse may lead to many adverse effects, including mental, physical and behavioral disorders as well as social and economic consequences. The World Health Organization provides estimates of current substance users worldwide. According to a recent study (WHO 2002),

- 2 billion people consume alcohol,
- 185 million people use illicit drug, and
- 1.3 billion people smoke worldwide.

Despite the negative impacts of substance abuse, the number of people dependent on drug is increasing every year at an alarming rate. The extent of harm due to a substance depends not only on how it has been abused but also on how harmful the substance is. In general, the adverse consequences of substance abuse can be classified into three categories (Nutt et al. 2007) as described below.

Addiction

Addiction may be regarded as a chronic brain disease. Repeated use of substances may cause compulsive drug seeking and use. According to WebMD (2016), " drugs contain chemicals that tap into the brain's communication system and disrupt the way nerve cells normally send, receive and process information". Such a change in the brain system reduces the ability of a person's self-control and may lead to long-term addiction. The combined effects of social, environmental and biological factors may then intensify drug-seeking behavior.

Health Effects

Substance abuse may lead to both physical and mental disorders. Long-term addiction to substances causes stain on the organs, venous and respiratory system. Physical effects of substance abuse include organ damage, hormone imbalance, cancer, prenatal and fertility issues, gastrointestinal disease and HIV/AIDS (NIDA 2012a). Chronic use of substances may also lead to neurological impairment, including depression, anxiety, memory loss, aggression, mood swings, paranoia and psychosis (NIDA 2012b). Alcohol addiction is also reported to be associated with prevalence of contagious diseases, liver problems, cancer, diabetes, and neuropsychiatric diseases (Rehm 2011).

Social and Economic Impacts

Substance abuse is often accompanied by adverse social and economic impacts (Sartor 1991). In fact, it has far-reaching effects on families, friends, employers, healthcare professionals and societies as a whole. Substance abuse may also contribute to violence, crime, financial problems, housing problems, homelessness and vagrancy (Sartor 1991). The estimate of the total overall cost of substance abuse in the United States exceeds $600 billion annually, which includes approximately $193 billion for illicit drugs, $193 billion for tobacco, and $235 billion for alcohol (WebMD 2016). According to a 2002 report, the estimate of the total societal cost of substance abuse in Canada is $39.8 billion (Rehm et al. 2006). It was also reported that legal substances (tobacco and alcohol) account for around 79.3% of the total cost. Health care cost in Canada has also increased in the last two decades.

1.1.2 Substance Abuse in Canada

The most commonly abused substances in Canada are alcohol, cannabis, cocaine, speed, salvia, methamphetamine and some particular pharmaceutical drugs (CADMUS 2012). Among these, alcohol and cannabis are the substances most frequently used by youth and young adults in Canada (CCSA 2007). According to a 2005 report, about 50% to 75% junior high and high school students are current users of alcohol in Canada (Patton et al. 2005). The alarming fact is that there is a tendency among the adolescents to consume an excessive amount of alcohol in a single occasion (binge drinking). In fact, more than 33% of junior high school students were reported to be binge drinkers in 2004 (Health Canada 2006). Unawareness of the consequences of alcohol abuse plays a major role in such behavior. According to a recent report (PHAC 2006), socioeconomic factors associated with alcohol and drug abuse include age, sex, income, education and cultural background.

In Canada, another widely used substance after alcohol is cannabis. According to a Health Canada survey, about 17% of the Grade 7-9 students are cannabis users (Health Canada 2006). Based on another survey reported by Adlaf et al. (2005), about 29% of 15-17 year olds and 50% of the 18-19 year olds , respectively, are lifetime cannabis users in Canada (Table 1.1). The startling fact is that about 3% to 5% of this population were daily users of cannabis (Adlaf and Paglia-Boak 2005, Poulin and Wilbur 2002, Poulin et al. 2005, Patton et al. 2005). In fact, Canadian teenagers are the top users of cannabis in the developed world (UNICEF 2013).

Table 1.1: Lifetime Cannabis Users in Canada.

Population	Percentage users
Grade 7-9	17%
15-17 year olds	29%
18-19 year olds	50%

Death at younger ages, impaired driving, chronic and acute health problems are some of the major consequences of substance abuse in Canada. As reported by the World Health Organization WHO (2016), there is a strong association between alcohol consumption and cancer. Cognitive and behavioral disorders are commonly observed among the cannabis users in Canada (CCSA 2015). Mental illness and impaired driving are also very common among them. In fact, concurrent disorder is increasing in Canada at an alarming rate, and has been recognized as one of the serious health issues for the Canadian youth population (CCSA 2009). In addition to health consequences, considerable amount of economic loss is entailed every year due to substance abuse in Canada. According to PHAC (2015), the total cost of alcohol consumption and other illicit substances were estimated to be $14.6 billion in 2002.

1.1.3 Statistical Modeling

Structural equation modeling (SEM) is a widely recognized and popular statistical technique in validating a hypothetical model about relationships among variables. It also provides a structure to analyze relationships between observed and latent variables, and allows causal inference. Its popularity has recently increased in many applications, including medical, health, biological, psychological, business, networking, economic and social sciences (Hayduk 1996). One of the main reason of increasing popularity of SEM is that it provides concise assessment of complex model involving many linear equations.

The basic concept of SEM was first introduced in late 1960s (Bollen and Noble 2011). Since then, extension of this technique and many other theoretical developments have been proposed in the literature, including SEM method for longitudinal data (Marsh and Yeung 1997), method for analyzing correlation among residuals to determine how correlation affects estimated parameters of the model (Brommera et al. 2014), and handling missing data (Allison 2003). Efficient computational algorithms and software development in recent decades have led to considerable interest in the application of SEM to complex modeling problems.

In general, SEM is a technique for multivariate data analysis, and involves a combination of two commonly used statistical techniques (Schreiber et al. 2006): factor analysis and regression analysis. Nowadays, many journals publish multivariate analysis of data using SEM (Hershberger 2003). The main purpose is to determine whether a hypothesized relationship or model is valid. In most cases, the model needs to be respecified based on the values of the goodness-of-fit criteria of the initially formulated model (Bollen and Noble 2011).

SEM can be an effective tool to depict relationships among health, substance abuse and the associated risk factors. One challenge of analyzing these type of data using SEM is that there could be many categorical variables in the data although SEM techniques for continuous variables are well documented in the literature (Muthen 2002), very few studies involving categorical data in SEM framework have been conducted so far. Note that SEM technique for continuous variables is not, in general, appropriate to analyze categorical data, as it may lead to erroneous conclusions (Johnson and Creech 1983). Recently, new methods have been developed to handle categorical data, including the limited information technique based on the weighted least squares method (Muthen (1981); see also Chapter 2). We consider this particular technique of SEM in our analysis, as there are some categorical variables in our data set (see Chapter 3).

1.2 Research Questions and Objectives

In Statistics, a construct or factor (also known as latent variable) reflects a continuum that is not directly measurable but could be summarized using different types of observed variables. An example is the construct of drug abusive environment, the various facets of which may be summarized using different types of observed variables, including cannabis ASSIST (Alcohol, Smoking and Substance Involvement Screening Test) scores, illicit drug use (yes/no), pharmaceutical drug abuse (yes/no) and harm from one's own drug use (yes/no). Another example is alcoholic environment, which could possibly be manifested by AUDIT (Alcohol Use Disorders Identification Test) scores, volume of alcohol drinking and whether or not the consumption is within the low-risk alcohol drinking guidelines. One more variable of relevance is health status, which might possibly be manifested by both observed and latent variables, such as alcohol and drug abusive environments (latent variables), and education, marital status and pharmaceutical drug abuse (observed variables). Note that there are other variables that might be associated with substance abuse and health status (e.g., age, sex, household size, smoking and cultural background), for which a regression analysis is useful. Based on the above discussion, the research questions of this study are presented as follows.

(a) What would be a reasonable hypothesis to explain causes of substance abusive behavior (i.e., cause and effect relationship)?

(b) What model would adequately describe the cause-and-effect relationship between the observed variables and health and substance-related latent variables?

(c) What covariates are significantly associated with alcohol and drug abusive environments and health status?

The main focus of this study is to address the above questions for the populations of two Canadian provinces, Saskatchewan and Manitoba. We will use data from the Canadian Drug Use and Monitoring Survey (CADMUS 2012), and consider socio-demographic and drug- and alcohol-related variables for analysis. In order to understand the complex mechanism of the cause-and-effect relationship, we will use SEM for statistical analyzes.

In Chapter 2, we describe the methodology of structural equation modeling with mathematical details, which include model specification, estimation and criteria for goodness of fit. The analysis of the data is then presented in Chapter 3. We conclude the thesis in Chapter 4 by presenting some additional considerations relevant to this work.

Chapter 2

Structural Equation Modeling

As mentioned in Chapter 1, we will use structural equation modeling to understand the cause-and-effect relationship among the variables that are related to substance abuse and health. In Section 2.1, we present the preliminary concepts of the structural equation modeling including variables and regression coefficients, notation, measurement model, structural model and structural regression model and steps of structural equation modeling. In Section 2.2, we describe model identification and in Section 2.3 we discuss model estimation. Goodness of fit criteria is discussed in Section 2.4. We conclude this chapter in Section 2.5 with a summary of the contents of this chapter.

2.1 Preliminary Concepts

As opposed to linear models where only observed variables are considered for analysis, SEM can account for both observed (also known as manifest) and latent (also known as factor) variables. This is a very unique feature of SEM. Another important feature of SEM is that it allows causal inference. The general idea is to formulate a model by hypothesizing interrelationships among the observed and latent variables via a path diagram, where the variables can serve as either correlates, predictors or consequences. In such a path diagram, a particular variable can not only serve as a predictor, but also be a response of other variables. SEM explicitly accounts for measurement errors in the indicators of a latent variable, as well as structural disturbances which typically reflect both omitted causes and measurement errors. We present some definitions and preliminary concepts of SEM in this section.

2.1.1 Variables and Regression Coefficients

Below we describe different types of variables used in SEM and distinguish between two types of regression coefficients, commonly known as path coefficient and factory loading.

- ***Exogenous Variable.*** Exogenous variables are similar to independent variables of linear models in concept. These variables do not receive a directional influence from any other variable, and may be considered as a causal input in the system. In a path diagram, the direction from an exogenous variable is always pointing out to another variable, indicated by a forward arrow $\rightarrow$.

- ***Endogenous Variable.*** An endogenous variable receives a directional influence from other variables (i.e., effects), and may also emit a directional influence to other endogenous variables (i.e., cause of other endogenous variables). An endogenous variable is subject to measurement errors, and therefore assumed to be influenced by a stochastic error term in modeling. This stochastic component accounts for influences of other known or unknown variables that are not considered in the analysis. In a path diagram, the direction is pointing in to endogenous variables (i.e., backward arrow $\leftarrow$), as well as sometimes have arrows pointing from them (i.e., forward arrow $\rightarrow$).

- ***Observed or Manifest Variable.*** An observed variable is a variable that can be observed and directly measurable. An observed variable can be either exogenous or endogenous.

- ***Latent Variable.*** A latent variable (also known as factor or construct) is a variable that is not directly observable or measured. It is assumed that a latent variable is manifested by other observed and/or latent variables. Note that a latent variable can be continuous or categorical and is assumed to be influenced by stochastic errors (also known as structural disturbances). A latent variable can be either exogenous or endogenous.

- ***Indicator.*** An observed variable used to measure a construct is called an indicator.

- ***Path Coefficient.*** The estimate of a direct effect from an observed variable to another variable is called a path coefficient, which can be interpreted simply as a regression coefficient.

- ***Factor Loading.*** The estimate of a direct effect from a latent variable to an indicator is called a factor loading, which can also be interpreted simply as a regression coefficient.

2.1.2 Notation

There is no standard set of notations for SEM. In this thesis, we adopt notations from Mulaik (2009), and present in Table 2.1. For path diagrams, we use standard symbols to describe different types of variables and directional effects: squares or rectangles to represent observed variables, circles or eclipses to represent latent variables, a line with a single arrowhead (→ or ←) to indicate directional effects of one variable to another variable, and a line (or curved line) with two arrowheads to indicate covariance between two variables.

Table 2.1: SEM notations, adopted from Mulaik (2009).

Variables and Coefficients	Notation
Exogenous latent variable	ξ
Endogenous latent variable	η
Endogenous observed variable	y
Exogenous observed variable	x
Number of latent endogenous variables	m
Number of latent exogenous variables	n
Number of observed endogenous variables	p
Number of observed exogenous variables	q
Path coefficient (exogenous to endogenous)	γ
Path coefficient (endogenous to endogenous)	α
Measurement error and structural disturbance	ϵ
Path coefficient (measurement error or structural disturbance to endogenous)	δ
Variance and covariance	ϕ

2.1.3 Measurement Model

Structural equation models have two components: a measurement model and a structural model. A measurement model involves hypotheses that relate the observed variables to the underlying latent variables, and the statistical techniques by which a measurement model is validated is called confirmatory factor analysis (CFA). As an example, Figure 2.1 displays a measurement model, where the exogenous latent variable ξ_1 is manifested by three endogenous observed variables y_1, y_2 and y_3, and the measurement errors in the endogenous observed variables are represented by ϵ_1, ϵ_2 and ϵ_3.

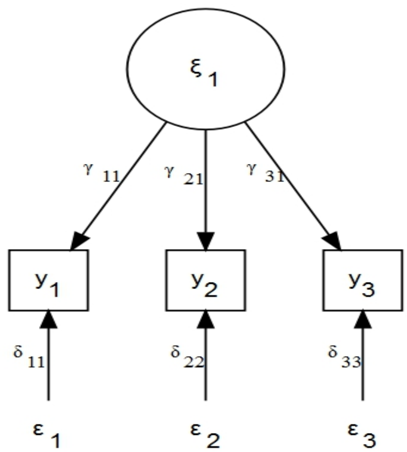

Figure 2.1: A measurement model with one endogenous latent variable ξ_1 and three exogenous observed variables y_1, y_2 and y_3; the observed variables are subject to measurement errors, modeled by ϵ_1, ϵ_2 and ϵ_3.

2.1.4 Structural Model and Structural Regression Model

A structural model involves hypotheses that relate the latent variables as posited by theory, and a structural regression model or SEM is a combination of a measurement model and a structural model. For example, the SEM in Figure 2.2 has both structural and measurement components. The structural part represents the hypothesis that the exogenous latent variables ξ_1 and ξ_2 each have direct effects on the endogenous latent variable η_1, and the measurement model hypothesizes three indicators (y_2, y_3 and y_4) for η_1, three indicators (y_5, y_6 and y_7) for ξ_1 and two indicators (y_8 and y_9) for ξ_2. Note that the structural disturbance is represented by ϵ_1, and the measurement errors by $\epsilon_2, \epsilon_3, \ldots, \epsilon_9$.

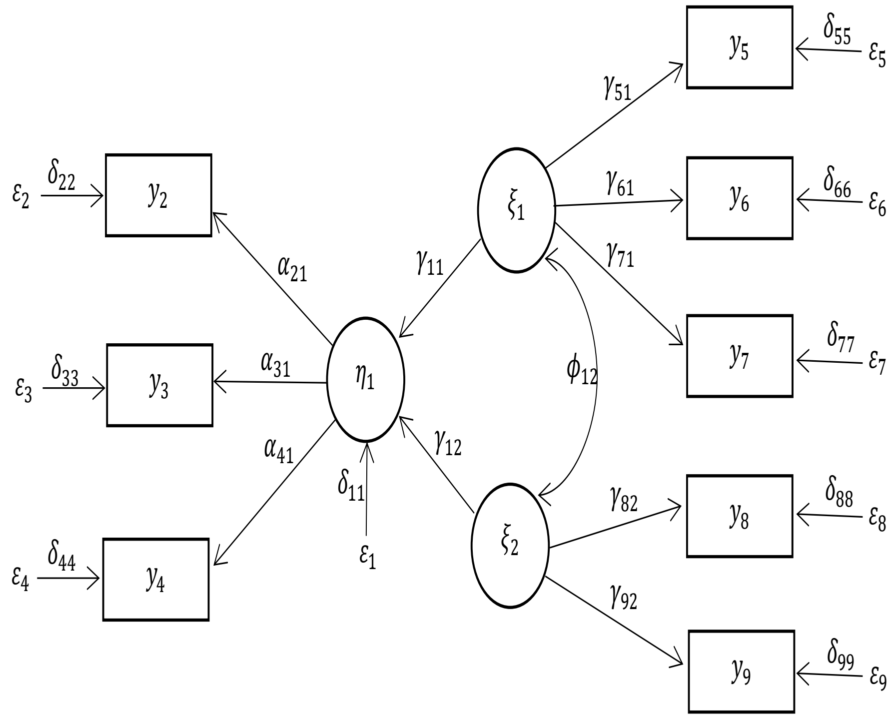

Figure 2.2: A structural regression model, consisting of a structural and a measurement component (Vij and Walker 2016). The structural part represents the hypothesis that the exogenous latent variables ξ_1 and ξ_2 each have direct effects on the endogenous latent variable η_1, and the measurement model hypothesizes three indicators (y_2, y_3 and y_4) for η_1, three indicators (y_5, y_6 and y_7) for ξ_1 and two indicators (y_8 and y_9) for ξ_2; the structural disturbance is represented by ϵ_1, and the measurement errors by $\epsilon_2, \epsilon_3, \ldots, \epsilon_9$.

2.1.5 Steps of Structural Equation Modeling

The SEM technique to validate a set of hypotheses involves four steps (Kline 2011) as described below.

Model Specification

Model specification involves specifying the relationship between observed and latent variables. The specified model is then represented by a path diagram. The specification of a model usually depends on knowledge of the context and previous research that has already taken place. One can initially specify more than one model for a particular study, and then compare the fits using goodness of fit criteria. Such a comparison may lead to a model that supports the data more adequately that any other models under consideration. If no model under consideration leads to a satisfactory fit, respecification of the hypothesis is necessary, as described below.

Model Estimation

After specifying a model, the next step is to estimate the underlying parameters that describe the proposed hypothesis. The estimation technique depends on the nature of the observed variables. Depending on the pattern of the data (continuous or categorical), a full estimation technique or a limited information estimation technique is employed (these methods are described in Section 2.3). In general, the estimation technique involves minimizing the difference between the sample covariance matrix and the implied covariance matrix, and requires an iterative procedure. The convergence is achieved if the absolute difference is less than a prespecified tolerance limit. Note that this algorithm is implemented in the "lavaan" package (Rosseel 2012) in R (R Core Team 2016).

Assessing Model Fit

The next step is to assess the fit of the model using goodness of fit criteria. There are many criteria available for SEM. In this thesis, we consider five criteria that are widely used to evaluate a structural equation model: chi-square goodness of fit, comparative fit index, Tucker-Lewis index, root mean square error of approximation and weighted mean square. Detail description of these criteria is presented in Section 2.4. If the values of the criteria lie within the recommended range (see Section 2.4), the proposed model is considered adequate in describing the data.

Respecification of the Model

A model requires a respecification if (1) it is not supported by the goodness of fit criteria, and/or (2) the estimates of the parameters are not interpretable. The most likely cause of (1) could be due to an untenable hypothesis specified by the model, whereas the most likely cause of (2) may be computational complexity (e.g., failure of convergence) rather than the model specification itself. Note that large residuals also indicate an unsatisfactory fit of the model.

In practice, it is common to have an unsatisfactory fit of the initially formulated model. In such cases, a respecification of the hypothesis is necessary. It is generally recommended that respecification should be based on the context of the problem and theoretical aspects, rather than on the values of the goodness of fit criteria (Bollen and Noble 2011).

In summary, the overall process could be very tedious and time consuming, as the hypothesis regarding a particular problem could be formulated in numerous ways.

2.2 Model Identification

An important consideration in SEM is model identification, or whether it is theoretically possible to obtain unique estimates of the parameters. The following description about model identification is taken from Kline (2011).

Necessary Conditions

There are two necessary but not sufficient conditions for model identification, as described below. *Necessary Condition 1.* Let there are r observed variables. Then the sample covariance matrix S has r(r+1)/2 nonredundant elements. The difference

$$r(r+1)/2 - (\text{number of model parameters})$$

is called the model degrees of freedom, and denoted by $\mathrm{df_M}$. A necessary condition for model identification is that $\mathrm{df_M}$ is nonnegative. A model for which $\mathrm{df_M} = 0$ is called a just-identified or saturated model, which is identified and always fit the data perfectly. A model for which $\mathrm{df_M} > 0$ is called an over-identified model, which is identified and has fewer parameters than the number of nonredundant elements in S. If $\mathrm{df_M} < 0$, the model is called under-identified; an under-identified model has more parameters than observations, and is not identified.

Necessary Condition 2. The second necessary condition is about scaling the latent variables. There are two types of latent variables in a structural equation model: error terms (disturbances for the structural component of the model and measurement errors for the measurement component of the model) and factors. The error terms are scaled using a technique called unit loading identification (ULI). Under this method, the path coefficient for a direct effect from an error is fixed to 1. For example, the unit loading identification constraint leads to $\delta_{11} = \delta_{22} = \ldots = \delta_{99} = 1$ for the model displayed in Figure 2.2. The error variances then need to be adjusted to account for this constraint. For example, if $\epsilon_2 \sim \mathrm{N}(0, 1)$ in Figure 2.2, then the equation $y_2 = \alpha_{21}\eta_1 + \delta_{22}\epsilon_2$ can simply be written as $y_2 = \alpha_{21}\eta_1 + \epsilon_2$ by taking $\delta_{22} = 1$ and $\epsilon_2 \sim \mathrm{N}(0, \delta_{22}^2)$. On the other hand, there are two approaches to scale the factors. The first approach is based on the ULI constraint, and involves fixing the loading for the direct effect from a factor to one of its indicators to 1. In general, the choice of the indicator for which the loading set to 1 is arbitrary. For example, one way to implement this approach for the model in Figure 2.2 is to take $\alpha_{21} = \gamma_{51} = \gamma_{82} = 1$. The second approach is called unit variance identification (UVI) constraint, by which the variance of a factor is fixed to 1. Under this method, all factor loadings are free parameters, and can be estimated uniquely provided that the other necessary and sufficient conditions are satisfied. Note that we have used the UVI constraint in our analyses in Chapter 3.

Sufficient Conditions

Suppose that the two necessary conditions are satisfied by a model. In addition, if all causal effects are unidirectional and the errors are uncorrelated, then the model is identified (Kline 2011). A model for which this additional condition is satisfied is called a recursive model. On the other hand, if a model has feedback loops and/or have correlated errors, then it is called a nonrecursive model. Sufficient conditions for a nonrecursive model to be identified involve extensive theoretical work. These conditions also depend on the type of the causal associations hypothesized via a model. We skip details about these conditions in this thesis. Interested readers may refer to Kline (2011). Note that the "lavaan (Rosseel 2012) package in R (R Core Team 2016) gives a warning message if a model appears to be unidentified.

2.3 Estimation

For individual i, let η_i and x_i be the vectors of latent endogenous and observed exogenous variables, respectively. The structural model can then be expressed as

$$\eta_i = \alpha + B\eta_i + \Gamma x_i + \epsilon_{1i} \tag{2.1}$$

where α is a vector of intercepts, B is a matrix of structural parameters (i.e, path coefficients for the latent variables), Γ is a matrix of coefficients for regressions of η_i on x_i, and ϵ_{1i} is a vector of disturbances. Letting y_i the vector of indicators (observed or latent) which manifest the latent variables η_i, the measurement model can be written as

$$y_i = \nu + A\eta_i + \epsilon_{2i} \tag{2.2}$$

where ν is a vector of intercepts, A is a matrix of factor loadings, and ϵ_{2i} is a vector of measurement errors. When y_i is observed continuous, least squares or maximum likelihood method can be used for estimation. For binary indicators, y_i is assumed to be latent, for which estimation is carried out by the so-called limited information method (Muthen 1981). A brief description of these two techniques are presented below.

Estimation when the Indicators are Continuous

The estimation is based on a comparison between the variance-covariance matrix of the observed variables and the sample covariance matrix S. In general, a discrepancy function of the form $F[\Sigma(\theta), S]$ is considered, where $\Sigma(\theta)$, the variance-covariance matrix of the observed variables, is expressed as a function of the model parameters θ (see Mulaik (2009) for theoretical detail of deriving an expression for $\Sigma(\theta)$). The discrepancy functions for ordinary least squares, maximum likelihood and generalized least squares are, respectively, (Mulaik 2009)

$$\frac{1}{2}tr[(S - \Sigma(\theta))'(S - \Sigma(\theta))], \tag{2.3}$$

$$\log|\Sigma(\theta)| + tr(S\Sigma[(\theta)]^{-1}) - \log|S| - p, \tag{2.4}$$

$$\frac{1}{2}tr[(\Sigma(\theta)S^{-1} - I)]^2. \tag{2.5}$$

The estimation is then carried out by minimizing the discrepancy function $F[\Sigma(\theta), S]$. This approach is sometimes referred to as full estimation technique, the application of which for categorical indicators may result in underestimation of the parameters (Distefano and Finney 2013). For this reason, the full estimation technique is generally not recommended if a factor is manifested by one or more categorical indicators.

Estimation when one or more Indicators are Binary

We assume that there is an underlying latent continuous variable associated with a binary indicator. For example, if the j^{th} indicator of individual i (say z_{ij}) is binary, then y_{ij} in (2.2) is assumed to be a latent continuous variable such that

$$z_{ij} = \begin{cases} 0 & \text{if } -\infty < y_{ij} \leq k_i, \\ 1 & \text{if } k_i < y_{ij} \leq \infty \end{cases}$$

where k_i is a threshold parameter defining the two categories of the binary indicator. Then, the model can be expressed using three components: a mean/threshold/reduced-form regression intercept structure, which includes the threshold parameters and denoted by σ_1; a reduced-form regression slope structure, denoted by σ_2; and a covariance/correlation structure, denoted by σ_3. The mathematical details of deriving these three components are given in Muthen (1981). The estimation is then carried out in three steps: (a) use logit of probit regressions to estimate σ_1 and σ_2 (denoted by s_1 and s_2, respectively), and tetrachoric correlations (pairwise correlations between the indicators) to estimate σ_3 (denoted by s_3), (b) find estimates of the covariance matrices of s_1, s_2 and s_3 using the maximum likelihood method, and (c) use the weighted least squares method for estimation, where the discrepancy function to be minimized is

$$F = \sum_{i} (s_i - \sigma_i)' W_i^{-1} (s_i - \sigma_i), \tag{2.6}$$

with W_i being the estimate of the covariance matrix for s_i. This three-step method is called the limited information technique, developed by Muthen (1981) and implemented in the "lavaan package (Rosseel 2012) of R (R Core Team 2016).

2.4 Goodness of fit criteria

Assessing the goodness of fit of the proposed model is an important step of structural equation modeling. Some widely used fit indices are briefly described below:

Chi-square Statistic

The chi-square statistic tests the hypothesis H_0 : the postulated model holds in the population, and all supporting assumptions also hold. This is a very popular method to evaluate a model. We can rewrite the null hypothesis as $H_o : \Sigma = \Sigma(\Theta)$; where Σ is population covariance, $\Sigma(\Theta)$ is sample or model implied covariance matrix. We can define the χ^2 statistic by (Schermelleh-Engel and Moosbrugger 2003)

$$\chi^2 = (N - 1)F[S, \Sigma(\hat{\Theta})] \tag{2.7}$$

where N is the sample size, S is the population covariance matrix and F is the discrepancy function. Researchers prefer non-significant χ^2 statistic. When χ^2 statistic is non significant the population covariance matrix is equal to sample covariance matrix, that indicates a good fit of the model.

Comaparative Fit Index (CFI)

CFI measures the relative improvement of the proposed model over that of the baseline model. This baseline model is also known as independent model where the observed variables are uncorrelated. Bentler (1990) proposed this index. CFI value ranges from 0 to 1. A CFI value ≥ 0.95 indicates a reasonable fit of the proposed model. Bentler (1990) defined CFI as

$$\text{CFI} = 1 - \frac{max[(\chi^2_{Ho} - df_{Ho}), 0]}{max[(\chi^2_{Ho} - df_{Ho})(\chi^2_b), df_b), 0]}, \tag{2.8}$$

where df_b and df_{Ho} are degrees of freedom for baseline and hypothesized model respectively.

Tucker-Lewis Index(TLI)

TLI is another incremental fit index. It is a measure of the discrepancy between the chi-square value of the hypothesized model and the chi-square value of the baseline model. Tucker and Lewis (1973) proposed this index and Bentler and Bonnet (1980) modified this index. Similar to CFI, minimum value of TLI is 0 and maximum value is 1. A TLI value ≥ 0.95 indicates an adequate fit of the proposed model. This index can be defined by

$$\text{TLI} = \frac{\chi^2_b/df_b - \chi^2_{Ho}/df_{Ho}}{(\chi^2_b/df_b) - 1}. \tag{2.9}$$

Root Mean Square Error of Approximation (RMSEA)

RMSEA measures the overall fit of the proposed model. It can be defined as (Schermelleh-Engel and Moosbrugger 2003)

$$\text{RMSE}A = \sqrt{max\left[\left(\frac{2F(\hat{\theta})}{d_{Ho}} - \frac{1}{N}\right), 0\right]}, \tag{2.10}$$

where d_{Ho} is the degrees of freedom of the model, $F(\hat{\theta})$ is the minimum of $F(\theta)$ and N is the sample size. We can use (2.10) for continuous outcome only. For categorical outcome, RMSEA is modified by replacing d_{Ho} by s_d, where s_d is the sample variance.

$$RMSEA = \sqrt{max\left[\left(\frac{2F(\hat{\theta})}{s_d} - \frac{1}{N}\right), 0\right]}. \tag{2.11}$$

Unlike other fit indices, RMSEA can be computed with confidence interval since the distribution of it is known (noncentral chi-square distribution). Recommended cut-off value of RMSEA is ≤ 0.05.

Weighted Root Mean Square Residual

WRMR is another measure to assess the overall fit of the proposed model. It was introduced by Muthen and Muthen (2001). WRMR can be used for continuous variables as well as for categorical variables. We can define WRMR for continuous variables as

$$\mathrm{WRMR} = \sqrt{\sum_{r}^{e} \frac{(s_r - \hat{\sigma}_r)^2 / v_r}{e}}, \tag{2.12}$$

where e is the number of sample statistic, s_r is the sample statistic and $\hat{\sigma}_r$ is the estimated statistic and v_r is the asymptotic variance or s_r. We can define it for categorical variable as

$$\mathrm{WRMR} = \sqrt{\frac{N \times F(\hat{\theta})}{e}}, \tag{2.13}$$

where $F(\hat{\theta})$ is the minimum of $F(\theta)$ and N is the sample size. The recommended cut-off value of WRMR is ≤ 0.90.

2.5 Summary

In this chapter, we have discussed model formulation and estimation for both categorical and continuous variable. In the next chapter, we will apply limited information estimation technique to the data and we will assess the goodness of fit of the model based on the recommended cut-off value that we have described in this chapter.

Chapter 3

Analysis

In this chapter, we present our hypotheses and results. A description of the data is presented in Section 3.1, followed by an exploratory data analysis (Section 3.2). In Section 3.3, we define the latent variables which are then subsequently used to formulate our hypotheses. In Section 3.4, we describe the rationale behind the proposed structural equation model for substance abuse, and then present model fits and results for Saskatchewan and Manitoba data. Our hypotheses and results to evaluate the effects of substance abuse on health are presented in Section 3.5. We conclude this chapter in Section 3.6 with a summary of the proposed models and results.

3.1 Data and Variables

The Canadian Alcohol Drug Use and Monitoring Survey (CADMUS) began in 2008 and continued until the end of 2012 (CADMUS 2012). Data on substance use and socio-demographic variables were collected annually under the direction of the Controlled Substances and Tobacco Directorate, Health Canada, and with collaboration from the Centre for Addiction and Mental Health (CAMH), the Centre for Addictions Research for BC (CARBC), Alberta Health Services, Manitoba Health and the Canadian Centre on Substance Abuse (CCSA). A telephone survey was conducted across all the provinces of Canada in 2012, with the target population being Canadians age 15 years or older (residents from the territories, permanent residents of institutions, people living in a household without a telephone and people with cell phones only were excluded from the survey). These resulted in a sample of size 11,090 in 2012.

Table 3.1: Definition of the observed variables under study.

Variable	Definition
Smoking	1 if smoker, 0 otherwise
Household size	Range: 1-7 for Saskatchewan and 1-11 for Manitoba
Age	1 if age $\geq$ 25, 0 if 15 $\leq$ age $<$ 25
Sex	1 if female, 0 otherwise
Cultural background	1 if not Caucasian, 0 if Caucasian
Education status	1 if completed high school/post secondary/university degree, 0 otherwise
Marital status	1 if married/living with partner, 0 otherwise
Pharmaceutical abuse in past 12 months	1 if used at least once, 0 otherwise
Alcohol Use Disorders Identification Test (AUDIT) score (Babor et al. 2001)	AUDIT assesses hazardous and harmful drinking; a score of 8 or more is conventionally regarded as hazardous or harmful.
Volume of alcohol consumed	Volume of alcohol consumed (in gram) per day
Low-risk drinking guideline (LRDG) for chronic effects (Butt et al. 2011)	1 if exceeds weekly limits for chronic effects, 0 otherwise
Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) score (WHO ASSIST Working Group 2002)	Ranges from 0 to 39; the higher the score, the greater the risk of experiencing problems
Illicit drug use in the past 12 months	1 if at least one illicit drug is used, 0 otherwise
Harm from ones own drug use in the past 12 months	1 if at least one harm, 0 if otherwise

We considered 2012 Saskatchewan and Manitoba data in this study, consisted of 1010 and 1009 individuals, respectively. Information on socio-economic, demographic and drug and alcohol related variables were reported in the survey. These include age, sex, education status, marital status, household size, cultural background, smoking, volume of alcohol drinking, alcohol AUDIT scores, Cannabis ASSIST scores, illicit drug use and pharmaceutical drug abuse. A description of these variables are presented in Table 3.1. Individuals with missing observation in any of the observed variables under study were removed from the analysis, resulted in samples of sizes 913 and 900 for Saskatchewan and Manitoba, respectively.

3.2 Exploratory Data Analysis

In this study, we consider ten binary variables (age, sex, cultural background, LRDG for chronic effects, Illicit drug use, pharmaceutical drug abuse, harm from one's own drug, education and marital status; see Table 3.1) and four continuous variables (Household size, AUDIT score, volume of alcohol consumed and ASSIST score). The distribution of the binary variables are presented in Table 3.2. Below we summarize the main points of interest.

- About 19% of the individuals were smokers (19.17% in Saskatchewan and 17.89% in Manitoba).
- Around 94% of the individuals were 25 years or older (93.43% in Saskatchewan and 94.45% in Manitoba).
- The study sample consisted of more females than males (64.07% females in Saskatchewan and 61.33% in Manitoba).
- The majority of the individuals were Caucasians (90.36% in Saskatchewan and 87% in Manitoba).
- About 11% of the individuals exceeded the low risk drinking guideline for chronic effects in Saskatchewan, whereas about 12% exceeded this guideline in Manitoba.
- About 7% in Saskatchewan and 8% in Manitoba used at least one illicit drug in the past 12 months.

- Only a very small fraction of individuals abused pharmaceutical drugs in the past 12 months (0.77% in Saskatchewan and 0.55% in Manitoba).
- About 1.5% of the individuals experienced at least one harm from drug use in the past 12 months.
- Around 85% of the individuals have completed high school or some post secondary/university degree in Saskatchewan and Manitoba.
- The majority of the individuals are married or living with their partner in both provinces (63.09% in Saskatchewan and 61% in Manitoba).

Table 3.2: Frequency distributions of the binary variables under study (n = 913 and 900 for Saskatchewan and Manitoba, respectively).

	Saskatchewan		Manitoba	
	Category		Category	
	0	1	0	1
Smoking	738 (80.83%)	175 (19.17%)	739 (82.11%)	161 (17.89%)
Age	60 (6.57%)	853 (80.83%)	50 (5.55%)	850 (94.45%)
Sex	585 (64.07%)	328 (35.93%)	552 (61.33%)	348 (38.67%)
Cultural background	825 (90.36%)	88 (9.64%)	783 (87%)	117 (13%)
LRDG for chronic effects	812 (88.94%)	101 (11.06%)	790 (87.78%)	110 (12.22%)
Illicit drug use	848(92.88%)	65 (7.12%)	822 (91.33%)	78 (8.67%)
Pharmaceutical abuse	906 (99.23%)	7 (0.77%)	895(99.45%)	5 (0.55%)
Harm	899 (98.47%)	14 (1.53%)	889 (98.78%)	11 (1.22%)
Education	143 (15.66%)	770 (84.34%)	127(14.11%)	773 (85.89%)
Marital status	337 (36.91%)	576 (63.09%)	351 (39.00%)	549 (61.00%)

Descriptive statistics (minimum, maximum, mean and standard deviation) for the numeric variables are presented in Table 3.5. We see that the distribution of household size is very similar for

Saskatchewan and Manitoba (mean ≈ 2.5 and SD ≈ 1.4). The average AUDIT score is slightly higher in Manitoba compared to Saskatchewan (mean = 2.942 and 3.877 in Saskatchewan and Manitoba, respectively), and these values support nonhazardous drinking in both Manitoba and Saskatchewan (a score of 8 or more is typically regarded as hazardous (Babor et al. 2001)). The average volume of drinking in the past 12 months is also slightly higher in Manitoba (average consumption is 2.036 gm per day in Saskatchewan and 3.397 gm per day in Manitoba). The average of the ASSIST scores are close in these two provinces (mean = 0.421 and 0.491 in Saskatchewan and Manitoba, respectively), and indicates low risk of experiencing health and other problems due to cannabis (a score between 0 and 3 refers to a pattern of use associated with a low risk of experiencing problems (WHO ASSIST Working Group 2002)).

Table 3.3: Summary measures for household size, AUDIT score, volume of alcohol drinking and ASSIST scores (n = 913 and 900 for Saskatchewan and Manitoba, respectively)

	Saskatchewan				Manitoba			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Household size	1	7	2.474	1.318	1	11	2.515	1.367
AUDIT score	0	29	2.942	3.413	1	33	3.877	3.374
Vol. of alcohol consumed	0	70	2.036	4.687	0	182	3.397	9.971
ASSIST score	0	26	0.421	2.101	0	25	0.491	2.150

3.3 Hypothetical Contrusts

We consider the observed variables and three hypothetical constructs (latent variables) to formulate our hypotheses. The latent variables are alcohol abusive environment, drug abusive environment and health (Table 3.4). In general, there is no set of rules to define a substance abusive environment, though many environmental and individual factors may play an important role, including genetic factors, peer pressure, depression and anxiety, mental illness, personality disorder, financial worries and family background. Such an environment may subsequently cause other people to abuse substance. In our analyses, we hypothesize that an alcoholic environment is manifested by (i.e., causal effect) variables related to alcohol abusive behavior (e.g., AUDIT score) and predicted by observed covariates. Similarly, a drug abusive environment is assumed to be manifested by variables related to drug abusive behavior and predicted by alcoholic environment and observed covariates. We also assume that health is manifested by some individual factors (e.g., education, marital status) and predicted by observed covariates. We first present our hypotheses and results for substance abuse (Section 3.4), and then consider structural equation modeling to evaluate cause-and-effect relationships between substance abuse and health (Section 3.5).

Table 3.4: Definition of the latent variables (hypothetical constructs) under study.

Variable	Definition
Alcohol abusive environment, a hypothetical construct	Manifested by alcohol abusive behavior and predicted by observed covariates
Drug abusive environment, a hypothetical construct	Manifested by drug abusive behavior and predicted by alcohol abusive environment and observed covariates
Health, a hypothetical construct	Manifested by education, marital status and pharmaceutical drug abuse and predicted by alcohol and drug abusive environments and observed covariates

3.4 Modeling Substance Abuse

For substance abuse, we present our hypotheses via the following three steps.

A Measurement Model for Alcohol Abuse

Alcohol abusive environment is taken as an endogenous latent variable (or factor). The indicator variables which are assumed to manifest alcohol abuse are AUDIT score, volume of alcohol consumption per day and LRDG. We hypothesize causal effects of alcohol abusive environment on the indicators, leading to a measurement model. This hypothesis explains the extent to which alcohol abusive environment is reflected in the measurements of the indicator variables. Then, five exogenous observed variables (smoking, household size, age, sex and cultural background) are hypothesized to predict abusive environment, for which a multiple regression analysis is considered. Figure 3.1 displays our hypothesis about alcohol abuse using a path diagram. Note that the structural disturbance for the endogenous variables is presented by ϵ_1 and measurement errors by ϵ_2, ϵ_3 and ϵ_4.

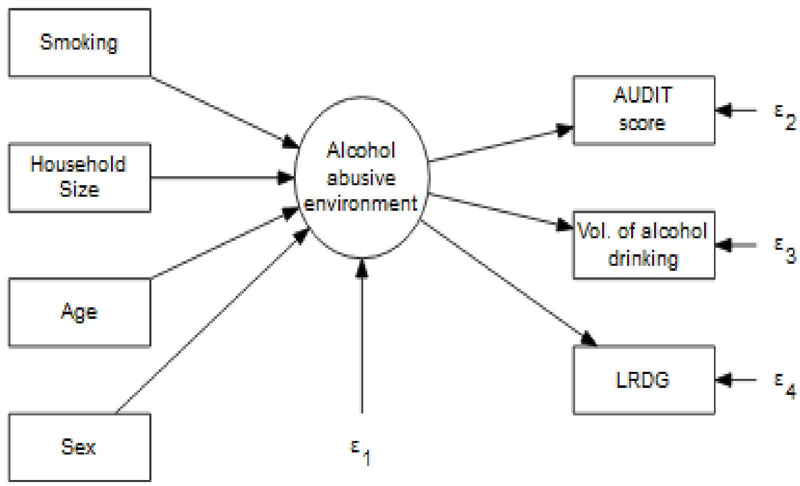

Figure 3.1: Measurement model for alcohol abusive environment: the latent variable (circle) is manifested by AUDIT score, volume of alcohol consumption and LRDG for chronic effects, and the covariates smoking, household size, age and sex are assumed to predict the latent variable.

A Measurement Model for Drug Abuse

Our hypothesis regarding drug abuse is similar to the measurement model for alcohol abuse. We consider drug abusive environment as an endogenous latent variable, and ASSIST score, illicit drug use, pharmaceutical drug abuse and harm from drug use as its indicators. Then, four exogenous observed variables (smoking, household size, age and sex) are considered for regression analysis (i.e., direct effects of these variables on drug abuse). The path diagram of this model is displayed in Figure 3.2 , where ϵ_2, ϵ_3 and ϵ_4 are measurement errors and ϵ_1 is the structural disturbance for the latent endogenous variable.

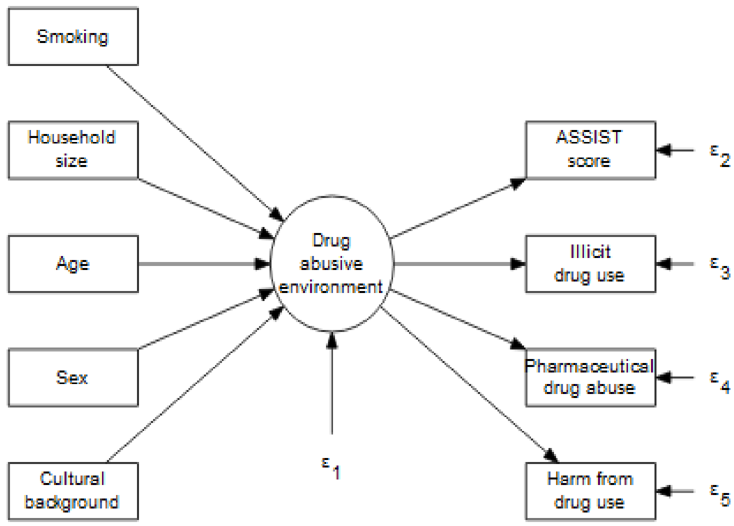

Figure 3.2: Measurement model for drug abusive environment : the latent variable (circle) is manifested by ASSIST score, illicit drug use, pharmaceutical drug abuse and harm from drug use, and the covariates smoking, household size, age, sex and cultural background are assumed to predict the latent variable.

A Structural Equation Model for Substance Abuse

The hypotheses regarding alcohol and drug abuse lead to our final model, where we also assume that alcohol abusive environment has a direct effect on drug abusive environment (i.e., alcohol abuse is hypothesized to affect drug abusive environment). This model is described in Figure 3.3 using a path diagram. In addition to our hypothesis about cause-and-effect relationships, we also assume that the following pairs of exogenous variables covary (i.e., measurement-error covariances): AUDIT score and volume of alcohol consumption, illicit drug use and volume of alcohol consumption, pharmaceutical drug abuse and harm from drug use, ASSIST score and pharmaceutical drug abuse, ASSIST score and harm from drug use, illicit drug use and pharmaceutical drug abuse and, illicit drug use and harm from one's own drug use.

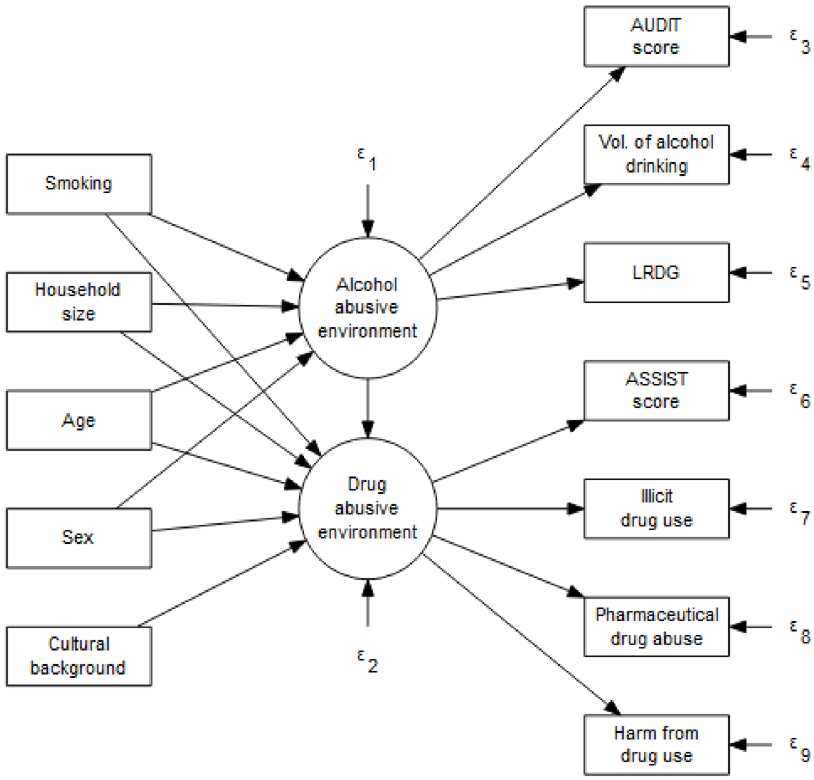

Figure 3.3: Structural model for substance abuse , which combines our hypotheses displayed in Figures 3.1 and 3.2.

We have considered linear models for the numeric variables, whereas logit links for the binary (categorical) variables. As described in Chapter 2, the weighted least squares method is used to fit the model, which is implemented in the lavaan package (Rosseel 2012) in R (R Core Team 2016).

3.4.1 Results for Saskatchewan

We fit the structural model of Figure 3.3 to Saskatchewan data using the unit variance identification constraint (see section 2.2). Values of the goodness of fit criteria are presented in Table 3.6. A nonsignificant chi-square (p-value = 0.519) supports our hypothesis, suggesting that the proposed model is consistent with the observed data. A CFI value of 0.997 (> 0.95) for the proposed model also supports our hypothesis. TLI, another incremental fit index which is a measure of the discrepancy between the chi-square value of the hypothesized model and the chi-square value of the baseline model, indicates an adequate fit of the proposed model (TLI = 0.996). Values of RMSEA = 0.017 (90% confidence interval ranges from 0.000 to 0.031) and WRMR = 0.685 also support our hypothesis. Overall, the proposed model regarding substance abuse leads to a satisfactory fit to Saskatchewan data.

Table 3.5: Goodness-of-fit indices for the proposed model (Figure 3.3) in analyzing Saskatchewan data.

Criterion	Recommended range	Saskatchewan
χ^2 statistic	-	30.967
p-value	> 0.05	0.519
Comparative Fit Index (CFI)	≥ 0.95	0.997
Tucker-Lewis Index (TLI)	≥ 0.95	0.996
Root Mean Square Error of Approximation (RMSEA)	≤ 0.05	0.017
Weighted Root Mean Square Residual (WRMR)	≤ 0.90	0.685

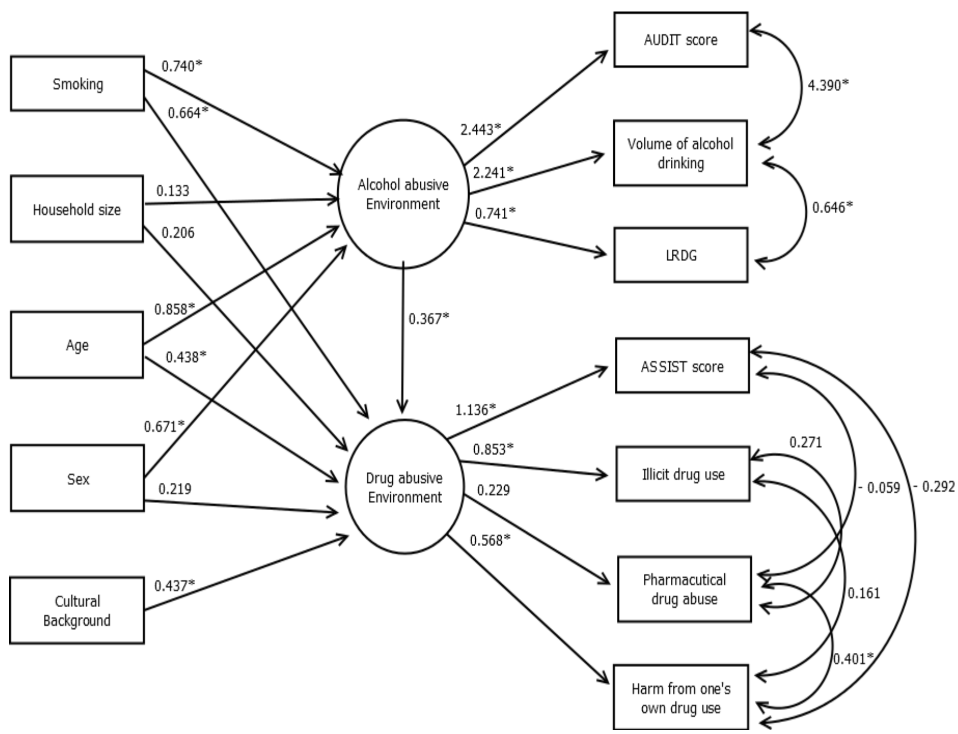

Figure 3.4: The hypothesized model for substance abuse along with estimates of the path coefficients and factor loadings (Saskatchewan data), an asterisk indicates statistical significance.

Table 3.6: Estimates and standard errors of the parameters for the proposed model (Figure 3.4) fitted to Saskatchewan data.

	Estimate	S.E.	p-value
Latent analysis for drug abusive environment			
ASSIST score	1.136	0.062	< 0.01
Illicit drug use	0.853	0.035	< 0.01
Phar. abuse	0.229	0.209	0.274
Harm	0.568	0.231	0.014
Latent analysis for alcohol abusive environment			
AUDIT score	2.443	0.142	< 0.01
Vol. of alcohol drinking	2.241	0.175	< 0.01
LRDG	0.741	0.046	< 0.01
Regression analysis for drug abusive environment			
Alcohol abusive environment	0.367	0.051	< 0.01
Smoking	0.664	0.139	< 0.01
Household Size	0.206	0.159	0.195
Age	0.438	0.175	0.012
Sex	0.219	0.135	0.104
Cultural Background	0.437	0.169	0.010
Regression analysis for alcohol abusive environment			
Smoking	0.740	0.118	< 0.01
Household Size	0.133	0.150	0.377
Age	0.858	0.151	< 0.01
Sex	0.671	0.102	< 0.01

Numerical results are presented in Table 3.7 (estimates of the parameters are also displayed in Figure 3.4). From latent analysis, we see that an alcoholic environment is significantly manifested by ASSIST score (p-value < 0.01), illicit drug use (p-value < 0.01) and harm from one's own drug (p-value = 0.013), whereas a drug abusive environment is significantly manifested by all its indicators (p-value is < 0.01 for each of ASSIST score, volume of alcohol drinking and LRDG for chronic effects). The factor loadings can be interpreted as regression coefficients. For example, the factor loading for AUDIT score is 2.443, suggesting that we expect a 2.443 point increase in AUDIT score given an increase of 1 point on alcoholic environment (a hypothetical construct).

Regression analysis (Table 3.7) reveals that alcoholic environment, smoking, age and cultural background are significantly associated with drug abusive environment, whereas smoking, age and sex are significantly associated with alcoholic environment. The estimates of the coefficients are all positive, suggesting a positive association of each of the covariates. For example, the estimate of the coefficient for smoking on alcoholic environment (a hypothetical construct) is 0.740, leading to the conclusion that we expect a 0.740 point increase in the latent variable for a transition from non smoking to smoking. Similarly, the estimate of the coefficient for cultural background on drug abusive environment is 0.437, suggesting that a drug abusive environment is more likely in the non Caucasian group compared to Caucasians. Also we see that there is a significant positive association between alcoholic and drug abusive environments (the estimate of the coefficient for alcoholism on drug abusive environment is 0.367).

Table 3.7: Goodness-of-fit indices for the proposed model (Figure 3.5) in analyzing Manitoba data.

Criterion	Recommended range	Manitoba
χ^2 statistic	-	29.254
p-value	> 0.05	0.253
Comparative Fit Index (CFI)	≥ 0.95	0.999
Tucker-Lewis Index (TLI)	≥ 0.95	0.999
Root Mean Square Error of Approximation (RMSEA)	≤ 0.05	0.014
Weighted Root Mean Square Residual (WRMR)	≤ 0.90	0.736

3.4.2 Results for Manitoba

We encountered convergence issues when applied the Saskatchewan model (Figure 3.4) to Manitoba data. Specifically, this model led to a negative estimate of the standard deviation for illicit drug use. Therefore, we considered a revised model to analyze the Manitoba data, hypothesizing that a drug abusive environment was manifested by ASSIST score, pharmaceutical drug abuse and harm from one's own drug use. This revised model is displayed in figure 3.5, along with the estimates of path coefficients and factor loadings. Table 3.8 gives the values of the goodness of fit criteria. The p-value for the chi-square test (p-value = 0.253) indicates that the proposed model is consistent with the observed data. The CFI and TLI values are 0.999 and 0.999, respectively, also suggesting an adequate fit of the proposed model. Moreover, values of RMSEA and WRMR are well within the recommended range for a satisfactory mode fit.

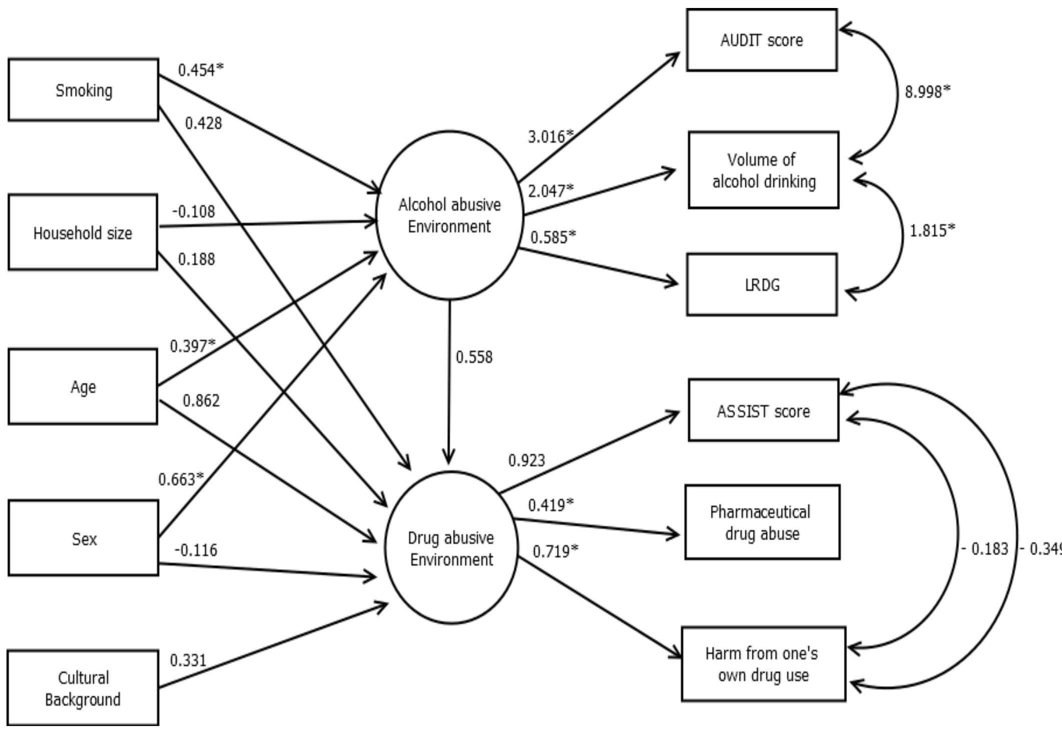

Figure 3.5: The hypothesized model for substance abuse along with estimates of the path coefficients and factor loadings (Manitoba data), an asterisk indicates statistical significance.

Numerical results for Manitoba data are presented in Table 3.9. We see that a drug abusive environment is significantly manifested by pharmaceutical drug abuse (p-value = 0.040) and harm from ones own drug use (p-value = 0.041). However, we found no significant causal association between a drug abusive environment and ASSIST score for Manitoba data (p-value = 0.057). On the other hand, an alcoholic environment is significantly manifested by all its indicators (p-value < 0.01 for each of AUDIT score, volume of alcohol drinking and LRDG). We also see positive causal association between each of the latent variables and their respective indicators. For example, the factor loading for AUDIT score is 3.016, suggesting that we expect a 3.016 point increase in AUDIT score given an increase of 1 point on alcoholic environment (a hypothetical construct).

Table 3.8: Estimates and standard errors of the parameters for the proposed model (Figure 3.5) fitted to Manitoba data.

	Estimate	S.E.	p-value
Latent analysis for drug abusive environment			
ASSIST score	0.923	0.485	0.057
Phar. abuse	0.419	0.204	0.040
Harm	0.571	0.279	0.041
Latent analysis for alcohol abusive environment			
AUDIT score	3.016	0.158	< 0.01
Vol. of alcohol drinking	2.047	0.314	< 0.01
LRDG	0.585	0.028	< 0.01
Regression analysis for drug abusive environment			
Alcohol abusive environment	0.558	0.298	0.061
Smoking	0.428	0.303	0.158
Household Size	0.188	0.265	0.477
Age	0.862	0.518	0.096
Sex	-0.116	0.191	0.542
Cultural Background	0.331	0.264	0.209
Regression analysis for alcohol abusive environment			
Smoking	0.454	0.093	< 0.01
Household Size	-0.108	0.154	0.481
Age	0.397	0.147	0.007
Sex	0.663	0.098	< 0.01

Regression analysis (Table 3.9) reveals that none of the predictors for drug abusive environment is statistically significant (an alcoholic environment might have a moderate positive effect as p-value = 0.061), whereas smoking, age and sex are significantly associated with an alcoholic environment. There is a positive association between each of these significant covariates and the latent variable. For example, the estimate of the coefficient for age on alcoholic environment is 0.397, suggesting that the adult population (age $\geq$ 25) is more likely to influence an alcoholic environment than does the younger population ($15 \leq$ age < 25).

3.5 Structural Equation Model for Health

A satisfactory model for substance abuse (Section 3.4) leads to the next stage of our analysis. The main goal is to develop a model for health, taking into account our hypotheses regarding substance abuse. Specifically, we consider an endogenous latent variable "health", with the assumption that the higher the value of this variable, the better the health status. We also assume that health is manifested by education, marital status and pharmaceutical abuse, and predicted by smoking, household size, age and sex (Figure 3.6).

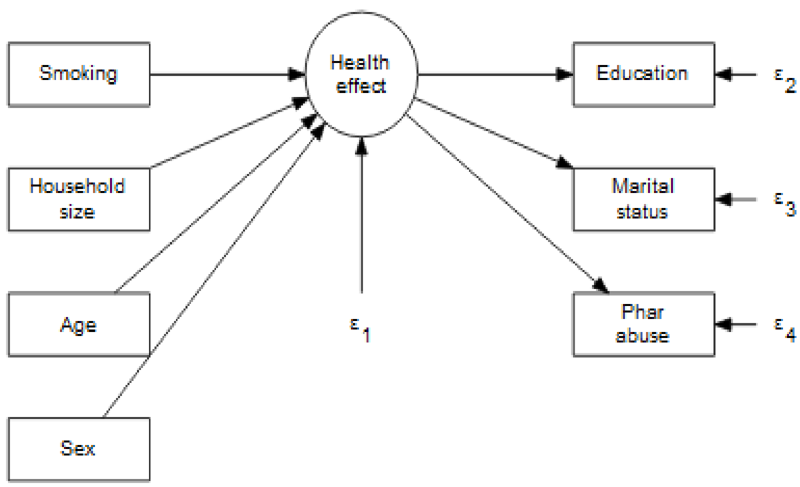

Figure 3.6: Measurement model for health: the latent variable (circle) is manifested by education, marital status and pharmaceutical drug abuse, and the covariates smoking, household size, age and sex are assumed to predict the latent variable.

3.5.1 Results for Saskatchewan

Our hypotheses regarding substance abuse and health are displayed in Figure 3.7, where alcohol and drug abusive environments are assumed to cause health problems. In our final model, we assume that alcohol abusive environment has a direct effect on drug abusive environment, and alcoholic and drug abusive environments have direct effect on health. In addition to our hypothesis about cause-and-effect relationships, we also assume that the following pairs of exogenous variables covary: ASSIST score and pharmaceutical drug abuse, ASSIST score and harm from one's own drug use, AUDIT score and volume of alcohol drinking, LRDG of chronic effects and volume of alcohol drinking, education and AUDIT score, education and volume of alcohol drinking, education and LRDG of chronic effects, education and ASSIST score, education and illicit drug use, marital status and pharmaceutical drug abuse, marital status and AUDIT score; and marital status and volume of alcohol drinking.

We fit the structural model of Figure 3.7 to Saskatchewan data. Values of the goodness of fit criteria are presented in Table 3.10: (a) a nonsignificant chi-square (p-value = 0.341) supports our hypothesis, suggesting that the proposed model is consistent with the observed data, (b) CFI and TLI values indicate an adequate fit of the proposed model, (c) values of RMSEA (0.009, with 90% confidence interval ranging from 0.000 to 0.025) and WRMR = 0.705 also indicate a reasonable fit of the proposed model.

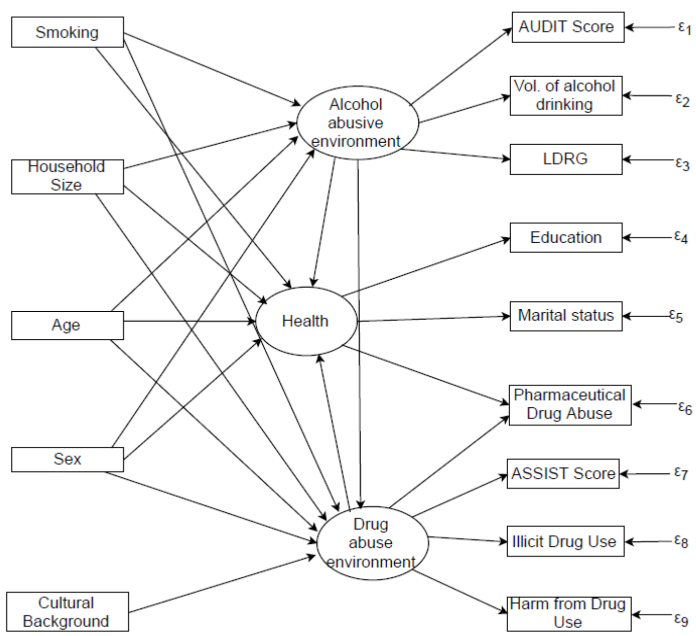

Figure 3.7: Structural equation model for health, combining our hypotheses displayed in Figures 3.3 and 3.6.

Table 3.9: Goodness-of-fit indices for the proposed model (Figure 3.8) in analyzing Saskatchewan data.

Criterion	Recommended range	Saskatchewan
χ^2 statistic	-	46.205
p-value	> 0.05	0.341
Comparative Fit Index (CFI)	≥ 0.95	0.999
Tucker-Lewis Index (TLI)	≥ 0.95	0.999
Root Mean Square Error of Approximation (RMSEA)	≤ 0.05	0.009
Weighted Root Mean Square Residual (WRMR)	≤ 0.90	0.705

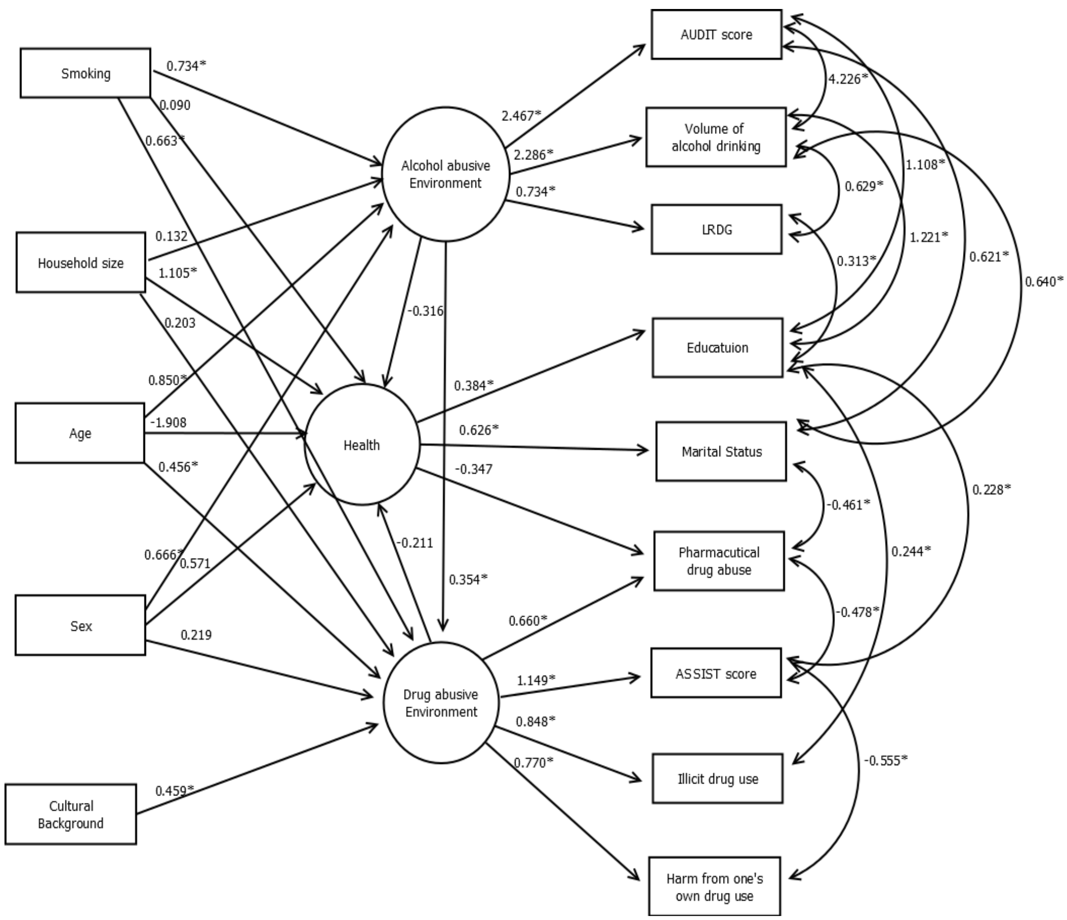

Figure 3.8: Fit of the hypothesized model for health (Saskatchewan data), along with estimates of the path coefficients and factor loadings (an asterisk indicates statistical significance).

Table 3.10: Estimates and standard errors of the parameters for the proposed model for health (Figure 3.8) fitted to Saskatchewan data.

	Estimate	S.E.	p-value
Latent analysis for drug abusive environment			
ASSIST score	1.149	0.060	< 0.01
Illicit drug use	0.848	0.035	< 0.01
Phar. abuse	0.660	0.128	< 0.01
Harm	0.770	0.077	< 0.01
Latent analysis for alcohol abusive environment			
AUDIT score	2.467	0.146	< 0.01
Vol. of alcohol drinking	2.286	0.181	< 0.01
LRDG	0.734	0.047	< 0.01
Latent analysis for health			
Education	0.384	0.062	< 0.01
Marital Status	0.626	0.117	< 0.01
Phar. abuse	0.347	0.226	0.125
Regression analysis for drug abusive environment			
Alcohol abusive environment	0.354	0.049	< 0.01
Smoking	0.663	0.135	< 0.01
Household Size	0.203	0.157	0.195
Age	0.456	0.172	< 0.01
Sex	0.219	0.133	0.099
Cultural Background	0.459	0.166	< 0.01
Regression analysis for alcohol abusive environment			
Smoking	0.734	0.118	< 0.01
Household Size	0.132	0.149	0.374
Age	0.850	0.151	< 0.01
Sex	0.666	0.102	< 0.01
Regression analysis for health			
Alcohol abusive environment	-0.316	0.196	0.108
Drug abusive environment	-0.211	0.109	0.052
Smoking	0.090	0.199	0.651
Household Size	1.105	0.405	< 0.01
Age	-1.908	0.386	< 0.01
Sex	0.571	0.214	< 0.01

Numerical results are presented in Table 3.11 (Estimates of the parameters are also displayed in Figure 3.8). Drug and alcohol abusive environments are significantly manifested by their respective indicators. We see that there is a positive association between each indicator and its factor. For example, one unit increase in the score of a drug abusive environment leads to an average increase of 1.149 units in ASSIST. Overall, the severity of an abusive environment increases with the increase of an indicator.

We see a significant impact of health on education and marital status (p-value < 0.01). The estimates of the factor loadings indicate positive associations: a healthy life is highly likely to cause (a) higher level of education (estimate of the factory loading for education is 0.384, where education is coded as 1 if high school or higher and 0 if less than high school), and (b) living together as a family (estimate of the factory loading for marital status is 0.626, where marital status is coded as 1 if married or living with a partner and 0 otherwise).

Let us turn now to regression analysis. We see (a) significant positive association between drug abusive environment and each of the following covariates: alcohol abusive environment, smoking, age, sex and cultural background, (b) significant positive association between alcohol abusive environment and each of the covariates smoking, age and sex, and (c) significant positive association between health and household size, health and sex, and significant negative association between health and age. Note that the association between health and drug abusive environment is marginally significant (p-value = 0.052). Some interesting findings of the regression analysis are presented below.

(i) There is a negative association between health and drug abusive environment (estimate of the coefficient is -0.211 with p-value = 0.052), suggesting that an increase in the severity of drug abusive environment worsen the health of individuals.

(ii) Alcohol abusive environment might have indirect effects on health: an alcohol abusive environment may lead to a drug abusive environment (estimate of the coefficient is 0.354 with p-value < 0.01), which in turn could have negative impacts on health as mentioned in (i).

(iii) We see significant positive associations of drug abusive and alcohol abusive environments, individually, with smoking (estimates of the coefficients are 0.663 and 0.734, respectively). Although smoking is not found to be significantly associated with health, it may lead to an environment that could worsen the health of individuals (i.e., indirect effects of smoking on health via a drug or alcohol abusive environment).

(iv) Females are more likely to experience negative health outcomes compared to males (estimate of the coefficient for sex on health is 0.571 with p-value < 0.01).

(v) The older population (age ≥ 25) is more likely to cause an abusive environment (estimates of the coefficients for age on drug and alcohol abusive enjoinments are 0.456 and 0.850, respectively), as well as they might experience more negative health outcomes compared to the younger population (estimate of the coefficient for age on health is -1.908).

(vi) Cultural background is highly likely to affect drug abusive environment (estimates of the factor loading of cultural background on health is 0.459 with p-value < 0.01). Cultural background may lead to drug abusive environment that could worsen the health of individuals (i.e., indirect effects of cultural background on health via a drug abusive environment).

3.5.2 Results for Manitoba

Similar to the model of Saskatchewan (Figure 3.4), we assume that alcohol abusive environment has a direct effect on drug abusive environment and alcoholic environment and drug abusive environment have direct effect on health for Manitoba also. This model is described in Figure 3.7 using a path diagram In addition to our hypothesis about cause-and-effect relationships, we also assume that the following pairs of exogenous variables covary: AUDIT score and volume of alcohol consumption; LRDG of chronic effects and volume of alcohol consumption; education and ASSIST score, illicit drug use, volume of alcohol consumption, AUDIT score and LRDG of chronic effects; marital status and ASSIST score, illicit drug use, pharmaceutical drug abuse, AUDIT score and volume of alcohol consumption.

Figure 3.9 displays the structural model for health. Table 3.8 represents the value of goodness of fit criteria. The null hypothesis for the chi-square test for this model is the same as previous model (Saskatchewan data). The p-value (0.090) indicates that the proposed model is consistent with the observed data. The CFI and TLI value of 0.998 (> 0.95) and 0.997 (> 0.95) indicate that the model appears to fit well. Values of RMSEA = 0.019 (90% confidence interval: (0.000, 0.033)) and WRMR = 0.759 also indicate a reasonable fit of our proposed model.

Table 3.11: Goodness-of-fit indices for the proposed model (Figure 3.9) in analyzing Manitoba data.

Criterion	Recommended range	Saskatchewan
χ^2 statistic	-	45.472
p-value	> 0.05	0.090
Comparative Fit Index (CFI)	≥ 0.95	0.998
Tucker-Lewis Index (TLI)	≥ 0.95	0.997
Root Mean Square Error of Approximation (RMSEA)	≤ 0.05	0.019
Weighted Root Mean Square Residual (WRMR)	≤ 0.90	0.759

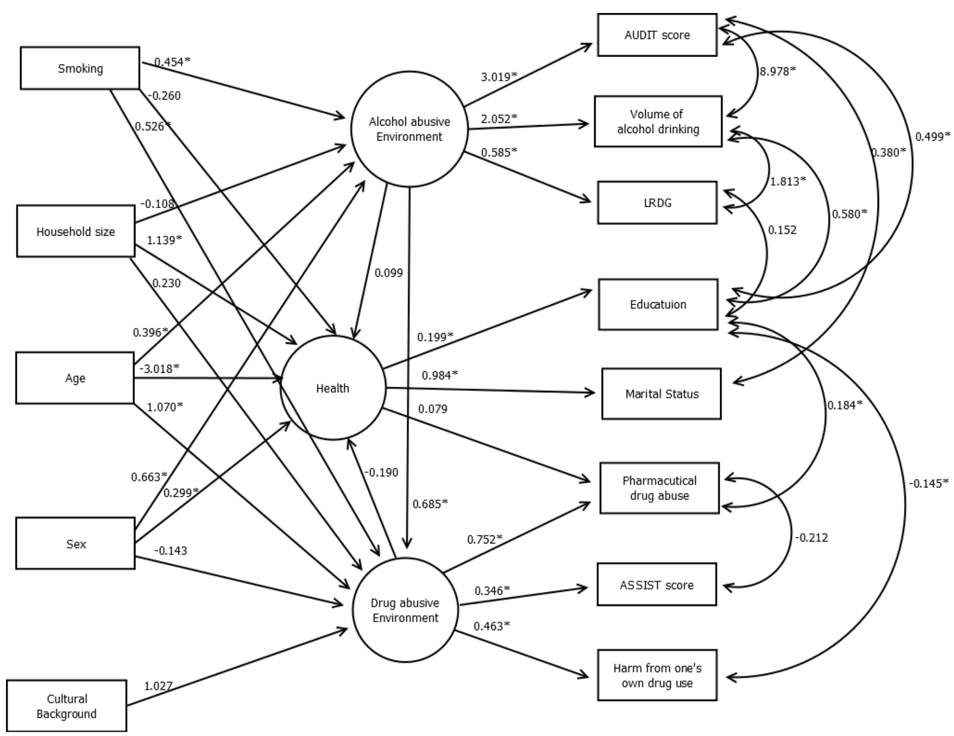

Figure 3.9: Fit of the hypothesized model for health (Manitoba data), along with estimates of the path coefficients and factor loadings (an asterisk indicates statistical significance).

Table 3.12: Estimates and standard errors of the parameters for the proposed model for health (Figure 3.9) fitted to Manitoba data.

	Estimate	S.E.	p-value
Latent analysis for drug abusive environment			
ASSIST score	0.752	0.091	< 0.01
Phar. abuse	0.346	0.056	< 0.01
Harm	0.463	0.044	< 0.01
Latent analysis for alcohol abusive environment			
AUDIT score	3.019	0.158	< 0.01
Vol. of alcohol drinking	2.052	0.313	< 0.01
LRDG	0.585	0.028	< 0.01
Latent analysis for health			
Education	0.199	0.050	< 0.01
Marital status	0.984	0.249	< 0.01
Phar. abuse	0.079	0.061	0.190
Regression analysis for drug abusive environment			
Alcohol abusive environment	0.685	0.099	< 0.01
Smoking	0.526	0.241	0.029
Household Size	0.230	0.307	0.454
Age	1.070	0.298	< 0.01
Sex	-0.143	0.228	0.529
Cultural Background	0.402	0.246	< 0.01
Regression analysis for alcohol abusive environment			
Smoking	0.454	0.093	< 0.01
Household Size	-0.108	0.153	0.481
Age	0.396	0.147	0.007
Sex	0.663	0.098	< 0.01
Regression analysis for health			
Alcohol abusive environment	0.099	0.122	0.418
Drug abusive environment	-0.190	0.100	0.058
Smoking	-0.260	0.149	0.081
Household Size	1.139	0.362	0.002
Age	-3.018	0.785	< 0.01
Sex	0.299	0.142	0.035

Numerical results are presented in Table 3.11 (Estimates of the parameters are also displayed in Figure 3.9). Latent analysis of drug and alcohol abusive environment indicate that both of these factors are significantly manifested by their indicators and these indicators and their respective factors are positively associated. For example, one unit increase in the score of a alcohol abusive environment leads to an average increase of 2.052 units in volume of alcohol drinking.

Education and marital status are significantly affected by health (p-value < 0.01). The estimates of the factor loadings indicate that both of these indicators are positively associated with health. The estimate of the factor loading for education is 0.199 which indicates that a healthy life is highly likely to cause higher level of education. The estimate of the factory loading for marital status (0.984) indicates that healthy life is highly likely to cause living together as a family.

The regression analysis indicate (a) significant positive association between drug abusive environment and each of the following covariates: alcohol abusive environment, smoking, age and cultural background, (b) significant positive association between alcohol abusive environment and each of the covariates: smoking, age and sex, and (c) significant positive association between health and household size, health and sex and significant negative association between health and age. Few important aspects of regression analysis are presented below.

(i) Health and drug abusive environment are negatively associated with p-value 0.058. This association is marginally significant because the p-value is very close to 0.05. The estimate of the factor loading for drug abusive environment on health is -0.190, which indicates that an increase in the severity of drug abusive environment worsens the health of individuals.

(ii) Household size and health are positively associated (estimates of the factor loading of household size on health is 1.139), which indicates that an increase unit increase in the household size can deteriorate the health condition of individuals.

(iii) Females are more likely to experience negative health outcomes compared to males (estimate of the coefficient for sex on health is 0.299 with p-value 0.035).

(iv) Age is positively associated with alcohol and drug abusive environment (estimates of the coefficients for age on drug and alcohol abusive enjoinments are 1.070 and 0.396, respectively) and negatively associated with health (estimate of the coefficient for age on health is -3.018), suggesting that the people who are 25 years or above are more likely to cause an abusive environment and they might experience more negative health outcomes compared to the younger population (age < 25).

3.6 Summary

We can summarize the following information:

(a) We can say that our hypothesized models for Saskatchewan and Manitoba sufficiently fit the data according to the fit criteria.

(b) We can also say that drug abusive environment have significant impact on ASSIST score, illicit drug use, pharmaceutical drug abuse and harm from one's own drug use; alcohol abusive environment have significant impact on AUDIT score, volume of alcohol drinking and LRDG for chronic effects; and health has significant effect on education and marital status for both Saskatchewan and Manitoba.

(c) Also, drug abusive environment is significantly affected by alcohol abusive environment, smoking, age and cultural background (for Manitoba only), whereas alcohol abusive environment is significantly affected by smoking, age and sex. We can also say that drug abusive environment, household size, age and sex have direct effect on health; and alcohol abusive environment, smoking and cultural background (for Manitoba) have indirect effect on health. One interesting fact is that though sex is significantly associated with drug abusive environment for both provinces, females are more likely to experience negative health effects compared to males in Saskatchewan, whereas males are more likely to experience negative health effects compared to females in Manitoba.

Chapter 4

Conclusion

Our first research question or objective was to find a reasonable model to explain substance abusive behavior. From the analysis of Saskatchewan and Manitoba, we can see that our hypothesized models for substance abuse (Figure 3.4 and Figure 3.5) for both provinces were adequate fit of the data. The additional information that we got from the analysis is that smoking, age, sex and cultural background are significantly associated with substance abusive environment (drug and alcohol abusive environment). Moreover, all the observed variables associated with alcohol abusive environment (AUDIT score, volume of alcohol drinking and LRDG for chronic effects) and drug abusive environment (ASSIST score, illicit drug use, pharmaceutical drug abuse and harm from one's own drug use) are statistically significant.

The second objective was to find model to describe the cause-and-effect relationship between observed variables and health and substance-related latent variables. From the analysis of Section 3.5, we can see that our proposed models (Figure 3.8 and Figure 3.9) adequately described the relationship. The analysis also indicates that latent variables - alcoholic and drug abusive environment also have direct effect on health.

The third research question was to find the covariates that are significantly associated with alcoholic and drug abusive environment and health. The significant covariates are represented in Figure 3.8 and Figure 3.9. The covariance between ASSIST score and pharmaceutical drug abuse, AUDIT score and volume of alcohol drinking, LRDG for chronic effects and volume of alcohol drinking, AUDIT score and education, education and volume of alcohol drinking, LRDG for chronic effects and education, pharmaceutical drug abuse and education, harm from one's own drug use and education; and AUDIT score and marital status are significant.

We found some similar studies that also involve finding an adequate model to describe the cause-and-effect relationship between observed variables and substance abuse. Wang et al. (2009) conducted a study to find the association among some socio-demographic variables and peer influence and parental knowledge on adolescent substance abuse. Their analysis indicates that peer influence and parental knowledge have association with substance abuse. Though our study does not include any information of peer influence, but the result of our study indicates that health is significantly manifested by education. Moreover, alcoholic and drug abusive environments have direct effect on health. So, we can say that there is an association between education and substance abuse. Also, their study indicates that Caucasian people are more likely to be substance users than non-Caucasian people, which agrees with the analysis of our model (Figure 3.8 and Figure 3.9). In our study, cultural background is significantly positively associated drug abusive environment.

There are numerous ways to specify a model. Given the same information, different researchers can formulate different models. There is no specific way to formulate a structural equation model. In our analysis, we had to hypothesize several models until we got a reasonable model according to the fit criteria. Our final model (Figure 3.8 and Figure 3.9) is an adequate fit the data based on the goodness of fit criteria. But it may not be the only model that sufficiently fit the data. There might be a better model than out hypothesized model for Saskatchewan and Manitoba. We assumed the same model for Alberta data also and it did not fit adequately indicating that same hypothesized model may not fit other data even if it fits similar data set. Another important limitation in our studies is that we discarded the missing values from the data assuming that the values are missing completely at random (MCAR).

This study was conducted for two provinces of Canada (Saskatchewan and Manitoba). In the future, we want to analyze the data of other provinces and compare the results of these provinces with the results of Saskatchewan and Manitoba.

References

Adlaf, E., Begin, P., and Sawka, E. (2005). Canadian addiction survey (cas): A national survey of canadians use of alcohol and other drugs: Prevalence of use and related harms: Detailed report. ottawa: Canadian centre on substance abuse. *Canadian Addiction Survey*. 4

Adlaf, E. M. and Paglia-Boak, A. (2005). Drug use among ontario students: Detailed OSDUS Findings, 1977-2005. Toronto: Centre for Addiction and Mental Health. *Centre for Addiction and Mental Health*. 4

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112(4):545–557. 5

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., and Monteiro, M. G. (2001). The alcohol use disorders identification test: Guidelines for use in primary care. *World Health Organization*. 22, 25

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, pages 238–246. 18, 19

Bentler, P. M. and Bonnet, D. C. (1980). Significance test and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3):588–606. 19

Bollen, K. A. and Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 108:15639–15646. 5, 6, 14

Bootzin, R. R. and Stevens, S. J. (2005). Adolescents, substance abuse, and the treatment of insomnia and daytime sleepiness. *Clinical Psychology Review*, pages 629–644. 1

Boyd, C., McCabe, S., and Teter, C. (2005). Medical and nonmedical use of prescription pain medication by youth in a detroit-area public school district. *Drug Alcohol Depend*, 81:37–45. 2

Brommera, J. E., Karellb, P., Aholad, K., and Karstinene, T. (2014). Residual correlations, and not individual properties, determine a nest defense boldness syndrome. *Behavioral Ecology*, 00(00):1–11. 5

Butt, P., Beirness, D., Stockwell, T., Gliksman, L., and Paradis, C. (2011). Alcohol and health in canada: A summary of evidence and guidelines for low-risk drinking. *Canadian Centre on Substance Abuse*. 22

CADMUS (2012). Canadian alcohol and drug use monitoring survey 2012: Microdata user guide. *Health Canada*. 4, 7, 21

CCSA (2007). *Substance Abuse in Canada: Youth in Focus*. Ottwwa, ON: Canadian Centre on Substance Abuse. 1, 4

CCSA (2009). Substance abuse in canada: concurrent disorders. *Ottawa, ON: Canadian Centre on Substance abuse*. 5

CCSA (2015). Substance abuse in canada: The effects of cannabis use during adolescence. *Canadian Centre on Drug Abuse*. 5

Distefano, C. and Finney, S. (2013). *Non-normal and categorical data in structural equation modeling*. Information Age Publishing. 17

Hayduk, L. (1996). *LISREL: Issues, Debates, and Strategies*. Baltimore, MD: Johns Hopkins University Press. 5

Health Canada (2006). Summary of results of the 2004-05 youth smoking survey. Ottawa: Health Canada. Retrieved from: https://uwaterloo.ca/canadian-student-tobacco-alcohol-drugs-survey/sites/ca.canadian-student-tobacco-alcohol-drugs-survey/files/uploads/files/CA04yss04_national_smoking_profile.pdf. 4

Hershberger, S. (2003). The growth of structural equation modeling: 1994-2001. *Structural Equation Modeling*, 10(1):35–46. 6

Johnson, D. R. and Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 48:398–407. 6

Kandel, D. and Faust, R. (1975). Sequence and stages in patterns of adolescent drug use. *Journal of the American Medical Asssociation*, 32(7):923–932. 2

Kline, R. B. (2011). *Principle and Practice of Structural Equation Modeling*. The Guilford Press. 12, 14, 15

Leyton, M. and Stewart, S. E. (2014). Substance abuse in canada: Childhood and adolescent pathways to substance disorder. *Ottowa, ON: Canadian Centre on Substance Abuse*. 2

Marsh, H. W. and Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psycholog*, 89(1):41–54. 5

Mulaik, S. A. (2009). *Linear Causal Modeling with Structural Equations*. CRC press. vii, 10, 16

Muthen, B. (1981). Factor analysis of dichotomous variables: American attitudes toward abortion. *Factor analysis and measurement in sociological research: A multi-dimensional perspective*. 6, 16, 17, 18

Muthen, B. O. (2002). Beyond sem: General latent variable modeling. *Behaviormetrika*, 29(1):81–117. 6

Muthen, L. and Muthen, B. (1998-2001). Exponentiated weibull family for analyzing bathtub failure-rate data. *Mplus Users Guide*. 20

NIDA (2007). Drugs, brain and behavior: The science of addiction. National Institute on Drug Abuse. Retrived from https://www.drugabuse.gov/publications/drugs-brains-behavior-science-addiction/preface. 1

NIDA (2012a). Medical consequences of drug abuse. Retrieved from: https://www.drugabuse.gov/related-topics/medical-consequences-drug-abuse. 3

NIDA (2012b). Methamphetamine: What are the long-term effects of methamphetamine abuse? Retrieved from: https://www.drugabuse.gov/publications/research-reports/methamphetamine/what-are-long-term-effects-methamphetamine-abuse. 3

Nutt, D., LA, K., Saulsbury, W., and Blakemore, C. (2007). Development of a rational scale to assess the harm of drugs of potential misuse. *Lancet*, 369(9566):1047–1053. 2

Patrick, M. E., Wightman, P., Schoeni, R. F., and Schulenberg, J. E. (2012). Socioeconomic status and substance use among young adults: A comparison across constructs and drugs. *Journal of Studies on Alcohol and Drugs*, 73(5):772–782. 2

Patton, D., Mackay, T., and Broszeit, B. (2005). Alcohol and other drug use by manitoba students. *Addictions Foundation of Manitoba*. 4

PHAC (2006). Social and economic factors that influence our health and contribute to health inequalitites.The State of Public Health in Canada. Retrieved from: http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2008/fr-rc/cphorsphc-respcacsp07a-eng.php. 4

PHAC (2015). The chief public health officer's report on the state of public health in canada: Alcohol consumption in canada. Public Health Agency of Canada. Retrieved from: http://healthycanadians.gc.ca/publications/department-ministere/state-public-health-alcohol-2015-etat-sante-publique-alcool/index-eng.php. 5

Poulin, C., Martin, D., and Murray, M. (2005). Newfoundland and labrador student drug use survey 2003: Summary report. St. Johns, NL: Department of Health and Community Services. *Department of Health and Community Services*. 4

Poulin, C. and Wilbur, B. (2002). Nova scotia student drug use survey 2002: Technical report. Halifax, NS: Nova Scotia Department of Health, Addiction Services and Dalhousie University. *Nova Scotia Department of Health*. 4

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 13, 15, 18, 29, 51

Rehm, J. (2011). The risks associated with alcohol use and alcoholism. *Alcohol Research & Health*, 34(2):135–143. 3

Rehm, J., Baliuna, D., Brochu, S., Fischer, B., Gnam, W., J, J. P., Popova, S., Sarnocinska-Hart, A., Taylor, B., Adlaf, E., Recel, M., and Single, E. (2006). The costs of substance abuse in canada 2002. *Canadian Centre on Substance Abuse (CCSA)*. 3

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36. 13, 15, 18, 29

Sartor, R. (1991). The social impact of drug abuse on community life. *Medicine and Law*, 10:205–208. 3

Schermelleh-Engel, K. and Moosbrugger, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2):23–74. 18, 19

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6):323–338. 5

Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38:1–10. 19

UNICEF (2013). Child well-being in rich countries: A comparative overview. Innocenti Report Card 11, UNICEF Office of Research, Florence. Retrieved from: `https://www.unicef-irc.org/publications/pdf/rc11_eng.pdf`. 4

Vij, A. and Walker, J. L. (2016). How, when and why integrated choice and latent variable models are latently useful. *Tranportation Research Part B: Methodological*, 90:192–217. viii, 12

Wang, J., Simons-Morton, B. G., Farhart, T., and Luk, J. W. (2009). Socio-demographic variability in adolescent substance use: Mediation by parents and peers. *Prevention Science*, 10(4):387–396. 46

WebMD (2016). Drug abuse, addiction and brain. Retrieved from: `http://www.webmd.com/mental-health/addiction/drug-abuse-addiction?page=2`. 3

WHO (2002). World health organization. Retrieved from: `http://www.who.int/substance_abuse/facts/global_burden/en/`. 2

WHO (2016). Management of substance abuse: Global status report on alcohol and health. World Health Organization. Retrieved from: `http://www.who.int/substance_abuse/publications/global_alcohol_report/en/`. 5

WHO ASSIST Working Group (2002). The alcohol, smoking and substance involvement screening test (assist): development, reliability and feasibility. *Addiction*, 97(9). 22, 25

Appendix

This chapter includes R (R Core Team 2016) programming codes used for structural equation model for health of Saskatchewan.

```
library(lavaan)
########################################################
full.dat<-read.csv("e://data/alcohol-full-data.csv",na.string=" ")
full.dat[1,]
#####################################################
sask.dat<-full.dat[full.dat$prov==8,]
sask.dat[1,]
nrow(sask.dat)
rm(full.dat)
###########################################################
table(sask.dat$age)
#age<-ifelse(sask.dat$age<=2,0,1)
age<-sask.dat$youth
table(age)
age1<-ifelse((sask.dat$age==3 | sask.dat$age==4),1,0)
table(age1)
age2<-ifelse(sask.dat$age>=5,1,0)
table(age2)
table(sask.dat$demo3)
#educ<-ifelse(sask.dat$demo3<=4,0,1)
educ<-sask.dat$educat4
```

```
table(educ)
educ1<-ifelse(educ==1,0,1)
table(educ1)
table(sask.dat$cultback)
cult<-ifelse(sask.dat$cultback==1,0,1)
sex<-ifelse(sask.dat$sex==2,0,1)
smoking<-sask.dat$smk12m
hhsize0<-sask.dat$demo2
hhsize0[hhsize0==98 | hhsize0==99]<-NA
hhsize<-ifelse(hhsize0<=4,0,1)
table(hhsize)
table(sask.dat$marstat3)
mstatus<-sask.dat$marstat3
mstatus1<-ifelse(mstatus==1,1,0)
table(mstatus1)
table(mstatus,educ)

########################################################
dat1<-data.frame(age=age,age1=age1,age2=age2,educ=educ,
educ1=educ1,cult=cult,sex=sex,
  smoking=smoking,hhsize=hhsize,mstatus=mstatus,mstatus1=mstatus1)

#########################################################
table(sask.dat$audit)
audit<-sask.dat$audit
alc.cons<-sask.dat$qfvolwk
lrdg<-sask.dat$lrdgchronic
#######################################################
```

```
dat2<-data.frame(audit=audit,alc.cons=alc.cons,lrdg=lrdg)
######################################################
assist<-sask.dat$asistcan

idrug<-ifelse((sask.dat$il6d12==1 | sask.dat$meth12m==1 |
sask.dat$inh12m==1 | sask.dat$other12m==1),1,0)

table(sask.dat$pharabuse)
table(sask.dat$pharhigh)
phar<-ifelse((sask.dat$pharabuse==1 | sask.dat$pharhigh==1),1,0)
table(phar)
table(sask.dat$dharm12m)
harm<-sask.dat$dharm12m
harm[harm==-999]<-NA
table(harm)
########################################################
dat3<-data.frame(assist=assist,idrug=idrug,phar=phar,harm=harm)
#######################################################
dat0<-data.frame(dat1,dat2,dat3)
nrow(dat0)

dat<-na.omit(dat0)
nrow(dat)
summary(dat)
dat<-data.frame(dat)
table(dat$educ1)
table(dat$mstatus1)
###########################################################
```

```
##########################################################
AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
                     '

fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
summary(fit, fit.measures = TRUE)

##########################################################
AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
           assist~~phar+harm
          audit~~alc.cons
          alc.cons~~lrdg
          educ1~~audit+alc.cons+lrdg
          educ1~~assist+idrug+phar+harm
```

```
          mstatus1~~assist+idrug+phar
          mstatus1~~audit+alc.cons
                    '

fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
summary(fit, fit.measures = TRUE)

##################################################
# Drop phar ~~ educ1

AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
           assist~~phar+harm
          audit~~alc.cons
          alc.cons~~lrdg
          educ1~~audit+alc.cons+lrdg
          educ1~~assist+idrug+harm
          mstatus1~~assist+idrug+phar
          mstatus1~~audit+alc.cons
                    '
```

```
fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
summary(fit, fit.measures = TRUE)

################################################
# Drop idrug ~~ mstatus1

AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
           assist~~phar+harm
          audit~~alc.cons
          alc.cons~~lrdg
          educ1~~audit+alc.cons+lrdg
          educ1~~assist+idrug+harm
          mstatus1~~assist+phar
          mstatus1~~audit+alc.cons
                      '

fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
```

```
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
summary(fit, fit.measures = TRUE)

##################################################
# Drop assist ~~ mstatus1

AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
           assist~~phar+harm
          audit~~alc.cons
          alc.cons~~lrdg
          educ1~~audit+alc.cons+lrdg
          educ1~~assist+idrug+harm
          mstatus1~~phar
          mstatus1~~audit+alc.cons
                     '

fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
```

```
summary(fit, fit.measures = TRUE)

################################################
# Drop harm ~~ educ1

AC.model<-'alc=~audit+alc.cons+lrdg
           drug=~assist+idrug+phar+harm
           health=~educ1+mstatus1+phar
           drug~alc+cult+smoking+hhsize+age+sex
           alc~smoking+hhsize+age+sex
           health~alc+drug+smoking+hhsize+age+sex
           assist~~phar+harm
          audit~~alc.cons
          alc.cons~~lrdg
          educ1~~audit+alc.cons+lrdg
          educ1~~assist+idrug
          mstatus1~~phar
          mstatus1~~audit+alc.cons
                     '
fit <- sem(AC.model, data = dat,ordered=c("lrdg",
       "idrug","phar","harm","educ1","mstatus1"),link="logit",
       std.lv=TRUE,std.ov=FALSE)

summary(fit, fit.measures = TRUE,estimates=FALSE)
summary(fit, fit.measures = TRUE)
########################################################
```