Principal component analysis of early alcohol, drug and tobacco use with major depressive disorder in US adults

Kesheng Wang[a,b,*], Ying Liu[a], Youssoufou Ouedraogo[a], Nianyang Wang[c], Xin Xie[d], Chun Xu[e], Xingguang Luo[b,f]

[a] Department of Biostatistics and Epidemiology, College of Public Health, East Tennessee State University, Johnson City, TN 37614, USA
[b] Center for Biological Psychiatry, Beijing HuiLongGuan Hospital, Peking University, Beijing, 100096, China
[c] Division of Biostatistics, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[d] Department of Economics and Finance, College of Business and Technology, East Tennessee State University, Johnson City, TN 37614, USA
[e] Department of Health and Biomedical Sciences, College of Health Affairs, University of Texas Rio Grande, Valley, Brownsville, TX 78520, USA
[f] Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06516, USA

ARTICLE INFO

ABSTRACT

Early alcohol, tobacco and drug use prior to 18 years old are comorbid and correlated. This study included 6239 adults with major depressive disorder (MDD) in the past year and 72,010 controls from the combined data of 2013 and 2014 National Survey on Drug Use and Health (NSDUH). To deal with multicollinearity existing among 17 variables related to early alcohol, tobacco and drug use prior to 18 years old, we used principal component analysis (PCA) to infer PC scores and then use weighted multiple logistic regression analyses to estimate the associations of potential factors and PC scores with MDD. The odds ratios (ORs) with 95% confidence intervals (CIs) were estimated. The overall prevalence of MDD was 6.7%. The first four PCs could explain 57% of the total variance. Weighted multiple logistic regression showed that PC_1 (a measure of psychotherapeutic drugs and illicit drugs other than marijuana use), PC_2 (a measure of cocaine and hallucinogens), PC_3 (a measure of early alcohol, cigarettes, and marijuana use), and PC_4 (a measure of cigar, smokeless tobacco use and illicit drugs use) revealed significant associations with MDD (OR = 1.12, 95% CI = 1.08–1.16, OR = 1.08, 95% CI = 1.04–1.12, OR = 1.13, 95% CI = 1.07–1.18, and OR = 1.15, 95% CI = 1.09–1.21, respectively). In conclusion, PCA can be used to reduce the indicators in complex survey data. Early alcohol, tobacco and drug use prior to 18 years old were found to be associated with increased odds of adult MDD.

1. Introduction

Major depressive disorder (MDD), also known as depression, is one of the most common mental disorders in the United States (US). It significantly affects a person's family and personal relationships, work or school life, sleeping and eating habits, and general health. Based on the data from the National Health and Nutrition Examination Survey (NHANES) from 2009 to 2012, 7.6% of Americans aged 12 and over had depression (classified as moderate or severe depressive symptoms in the past 2 weeks) (Pratt and Brody, 2014). The National Institute of Mental Health has estimated that 16.1 million adults (6.7% of the US adult population) aged 18 years or older had reported at least one major depressive episode in the past year (NIMH, 2015); while the economic burden of depression in the US was estimated to be $210.5 billion in 2010 (Greenberg et al., 2015).

Previous studies have shown that several behavioral factors have been associated with depression including alcohol use and alcohol use disorder (Gilman and Abraham, 2001; Hasin and Grant, 2002; Sullivan et al., 2005; Boden and Fergusson, 2011; Kinyanda et al., 2011; Mugisha et al., 2015), and smoking and nicotine dependence (Glassman et al., 1990; Breslau et al., 1991; Kendler et al., 1993; Husky et al., 2008). Furthermore, MDD is highly comorbid with substance use disorders (SUDs) (Merikangas et al., 1998; Grant et al., 2016; Han et al., 2017) including alcohol and marijuana abuse/dependence (Greenbaum et al., 1991; Grant and Harford, 1995; Miguel-Hidalgo et al., 2010), opioid abuse (Brooner et al., 1997), and nicotine dependence (Cardenas et al., 2002; Chou et al., 2016; Huang et al., 2017; Martínez-Ortega et al., 2017).

Few studies have focused on the early alcohol, smoking and drug use on MDD. For example, tobacco use in late adolescence may predict

MDD in the early 20s (Brook et al., 1998); while early onset of regular tobacco use is a predictor for lifetime drug use and depressive disorders (Hanna and Grant, 1999). Using a community-based longitudinal design, it has been shown that the cumulative frequency of drug use (alcohol, marijuana, and other illegal drugs) covering the period of childhood and early adolescence, middle and late adolescence increases the risk of later MDD, alcohol dependence, and SUDs; while marijuana use and the use of other illicit drugs increase the risk of later MDD, alcohol dependence, and SUDs (Brook et al., 2002). Another longitudinal data suggested that smoking increases the risk for MDD in women (Pasco et al., 2008). In addition, current smoking but not current smokeless tobacco use among adolescents and adults is associated with MDD using the NSDUH 2011 data (Redner et al., 2014). Another study showed that early-onset cannabis smokers had an increased risk of depression (Fairman and Anthony, 2012). Furthermore, adolescent alcohol use is positively associated with later depression in a population-based U.K. cohort (Edwards et al., 2014). Boden and Fergusson (2011) also concluded that increasing involvement with alcohol increases the risk of MDD. A recent study reported that the volume of drinking, age of first drink, binge drinking, and alcohol dependence were associated with MDD using the 2010–2012 National Survey on Drug Use and Health (NSDUH) data (Jetelina et al., 2016; Wagner, 2016).

The NSDUH is a nationally representative survey that is conducted annually to assess the prevalence and correlations of drug use in the US for the population aged 12 years and older. The 2013 and 2014 NSDUH data provided 17 early alcohol, tobacco and drug use variables prior to 18 years; whereas these variables maybe intercorrelated. Principal component analysis (PCA) can transform the set of correlated variables into a reduced number of uncorrelated variables known as principal components (PCs) or factors and then overcome the disturbance of the multicollinearity of the risk factors. PCA has been vastly used in social sciences, health service, and health sciences research (Lattin et al., 2002; Navarro et al., 2011; O'Rourke; Larry Hatcher, 2013; Jackson et al., 2015; Wang et al., 2017). Previous studies suggest that PCA is suitable for use with continuous and/or binary variables (Filmer and Pritchett, 2001; Howe and Hargreaves, 2008; Kolenikov and Angeles, 2009; O'Rourke and Larry Hatcher, 2013; Jackson et al., 2015). To our updated knowledge, this is the first study to examine the factor structure of early alcohol, drug and tobacco use prior to 18 years of age using PCA and the associations of these factors with MDD using the NSDUH data.

2. Materials and methods

2.1. Study population

Data were extracted from the 2013 and 2014 NSDUH data. The NSDUH is conducted by the Substance Abuse and Mental Health Services Administration of the United States Department of Health and Human Services. The NSDUH is a survey of the civilian non-institutionalized individuals (≥12 years old) in the U.S. to provide annual population estimates of substance use and health. By employing a multistage area probability sampling of the 50 states and the District of Columbia, we obtain a representative sample of all civilian non-institutionalized individuals who are ≥12 years old and response the survey items. The NSDUH measures the prevalence and correlations of drug use in the U.S. quarterly and annually. Information from the survey contains data on the use of illicit drugs (e.g., marijuana, cocaine, hallucinogens, heroin, and inhalants), alcohol and tobacco, and nonmedical use of prescription drugs (e.g., pain relievers, tranquilizers, and sedatives). The survey included questions from the Diagnostic and Statistical Manual (DSM) of Mental Disorders along with substance abuse treatment history and perceived need for treatment. Respondents in the age group of 12–17 years were also asked of any illegal activities, neighborhood environment, and drug use by friends, etc. The background information of the

respondents included gender, race/ethnicity, age, marital status, educational level, job and veteran statuses, and current household composition. Details of the survey design and data collection methods are published elsewhere (Center for Behavioral Health Statistics and Quality, 2014, 2015). The total sample size of the combined 2013–2014 NSDUH data is 110,431. The current analysis was restricted to participants aged 18 years and older. There was an Institutional Review Board exemption due to secondary data analysis.

2.2. Measurements

2.2.1. Major depressive disorder

MDD in the NSDUH 2013–20414 data was determined based on the DSM-IV (American Psychiatric Association, 1994). Respondents were classified as experiencing MDD in the past year if they (a) reported meeting at least five of nine criteria within a two-week period at any point in their lifetime (i.e., depressed mood, diminished pleasure, weight gain/loss, fatigue, sleep changes, psychomotor distress, feelings of worthlessness, difficulty concentrating, or recurrent thoughts of suicide/death), one of which must have been either depressed mood or diminished pleasure, and (b) answered affirmatively to a question that queried whether they felt a depressed mood or diminished pleasure (in addition to other symptoms the respondent affirmed) for at least two weeks within the past year. A total of 6,239 adults with MDD and 72,010 controls were selected from the combined 2013 and 2014 NSDUH data.

2.2.2. Social-demographic factors

Gender was self-reported as male or female. Age was classified into four categories: 18–25 years, 26–49 years, 50–64 years and 65+ years. Race consisted of five subgroups: Whites, African American (AA), Asian, Hispanic, and Other. There were three categories of marriage status: married; widowed/divorced/separated, and never been married. The four categories of annual income were: < $20,000, $20,000-$49,999, $50,000-$74,999 and $75,000 or more. Health status had 4 levels: excellent, very good, good and fair/poor. Educational level was classified into less than high school, high school, some college and college graduate. There were three levels of county type: large metro, small metro and non-metro according to the revised definitions of metropolitan statistical areas (MSAs) by the Office of Management and Budget (OMB) (Office of Management and Budget, 2003, Bulletin No. 03–04). Large metropolitan areas were defined to have a population of at least 1 million, small metropolitan areas have a population fewer than 1 million, and non-metropolitan areas fell outside MSAs.

2.2.3. Substance use disorders in the past year

Nicotine dependence was defined as the average score over 17 questions pertaining to five aspects of dependence which based on the dependence criteria according to the Nicotine Dependence Syndrome Scale (NDSS). This average score was only computed for those respondents who answered all 17 questions (Shiffman et al., 2004). Based on the NDSS score, a respondent was defined as having nicotine dependence if their average score was greater than or equal to 2.75. Illicit drug or alcohol abuse or dependence in the past year was defined by being dependent on any illicit drug or alcohol in the past year or illicit drug or alcohol abuse in the past year. Illicit drug abuse is defined as abusing any of the following substances: marijuana, hallucinogens, inhalants, tranquilizers, cocaine, heroin, pain relievers, stimulants, or sedatives. Illicit drug dependence was classified as being dependent on any of these following substances: marijuana, hallucinogens, inhalants, tranquilizers, cocaine, heroin, pain relievers, stimulants, or sedatives.

2.2.4. Early alcohol, drug and tobacco use prior to 18 years old

Seventeen variables about early alcohol, drug and tobacco use are listed in Table 2. These variables were defined by "first used the drug prior to age 18", each of them had 2 levels (yes or no).