Fall 12-15-2017

Multivariate Examination Of Risk Factors Associated With Chronic Hepatitis C Virus Infection In The United States Using National Health And Nutrition Examination Survey From 2003 to 2014

Vaishnavi Pattabiraman

Follow this and additional works at: https://scholarworks.gsu.edu/iph_theses

MULTIVARIATE EXAMINATION OF RISK FACTORS ASSOCIATED WITH CHRONIC HEPATITIS C VIRUS INFECTION IN THE UNITED STATES USING NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY FROM 2003 TO 2014

by

VAISHNAVI PATTABIRAMAN

M.Sc., BHARATHIAR UNIVERSITY, COIMBATORE, INDIA

M.S., EMORY UNIVERSITY, ATLANTA, GA 30329

A Thesis Submitted to the Graduate Faculty of Georgia State University in Partial Fulfillment of the Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GA 30303

2017

MULTIVARIATE EXAMINATION OF RISK FACTORS ASSOCIATED WITH CHRONIC HEPATITIS C VIRUS INFECTION IN THE UNITED STATES USING NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY FROM 2003 TO 2014

By:

Vaishnavi Pattabiraman

Approved:

_____

Dora Il'yasova, PhD

Committee Chair

_____

William Thompson, PhD

Committee Member

_12/01/2017_____

Date

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Vaishnavi Pattabiraman

Signature of Author

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

supervisor(s) by allowing me to complete my school related deadlines on time. Michelle, Gerry and Julie are my friends at work who have cheered me all along this journey. Grateful to Betsy Gunnels who is not only an excellent part-time faculty at the SPH while working as a senior level statistician at CDC but has been very open in continuing to stay connected with me. She is the reason for connecting me with Dr. William Thompson (Bill), who is my SAS Guru.

I am short of words in expressing my sincerest thanks to Bill. Without Bill's generous support, I would not have completed my practicum or my thesis. My connection with Bill happened because of God's will through Betsy and I give thanks to God for this. I have learned so much from my year and a half interaction with Bill. I have personally not experienced anybody at a senior level position who have mentored students with genuine interest and care like Bill. I knew almost nothing about SAS when I started my practicum work with Bill. However, he went above and beyond what is expected of him in teaching me SAS over Skype business calls (again thanks to CDC for this) from scratch. He has trained me to think like an epidemiologist while being open to my ideas as a lab scientist. He readily agreed to serve on my thesis committee when I approached him, so much thanks to him for all the support in the past, present and in future.

Immensely thankful to Dr. Dora Il'yasova, Associate Professor at the SPH for accepting to serve as my committee chair. Dora was graceful in accepting the project that I already had going with Bill and offered excellent ideas based on her expertise to develop this as a stronger project. My committee has been very flexible all along and I am vastly thankful for the same. Flexibility and prompt responses from the committee was the key to complete this work on time.

Much thanks to the SPH's graduate advisors - Gina Sample and Jessica Pratt who were always ready to answer questions and helped facilitate my smooth sailing through this program. Their tremendous commitment and support to the students in the program has only made my life easier as a part-time student. Finally, heartfelt thanks to the SPH program and faculty members for developing a curriculum suitable for full time working professionals. None of this would have been possible without your flexible graduate program.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Introduction:

Hepatitis C Virus (HCV) Infection is the most common blood-borne infection in the world with a global prevalence of ~3%. It also represents an underestimated and under-recognized viral infection because it is asymptomatic during the initial period of infection, which tends to span several decades. However, after establishing itself as a chronic state, HCV infection often leads to severe debilitating liver conditions such as cirrhosis, hepatocellular carcinoma to name a few resulting in poor quality of life, increased healthcare costs and mortality. Several earlier studies have examined risk factors associated with HCV. However, a comprehensive study that has simultaneously evaluated a wide range of addictive risk behaviors associated with HCV has not been conducted to date. This type of investigation will help to identify at-risk populations for HCV and provide valuable information regarding how one might efficiently link them to appropriate treatment and care.

Aim:

The primary aims of this study were 1). To estimate chronic HCV infection (CHI) prevalence in non-institutionalized U.S. adult population from 2003-2014 2). To perform a multivariate examination of all known behavioral risk factors significantly associated with CHI and 3). To identify less invasive questions regarding risk behaviors associated with CHI that could be used to predict state-level CHI prevalence using other state-specific data sources such as The Behavioral Risk Factor Surveillance System (BRFSS).

Methods:

The study utilized nationally representative data from National Health and Nutrition Examination Survey (NHANES) for the years 2003-2014. HCV RNA positive persons were CHI positive population. Bivariate analyses were performed to examine the frequency distributions of the study's primary dependent variable (CHI) and all independent variables (demographical + behavioral risk factor variables). The analysis sample included 11,596 adults aged 20-59 years. Risk factors for CHI were examined using both bivariate and multivariate logistic regression analyses. We first conducted weighted bivariate logistic regression analyses to examine the

relationships of dependent and independent variables without controlling for potential confounders. We then conducted weighted multivariate adjusted logistic regression analyses to examine the relationships between the dependent and independent variables while controlling for potential confounders. Odds ratios (OR), 95% confidence limits (CL) and *p*-values were calculated. A *p*-value < 0.05 was considered statistically significant. SAS 9.4 was used for all statistical analyses.

Results:

The estimated number of CHI adults aged >/= 20 years in 2014 was 1.93 million leading to an estimated CHI prevalence of 0.7%. Injection drug users (IDU) had the highest CHI prevalence of 30.24% by bivariate analyses. Multivariate logistic regression analysis in the chronic HCV full model indicated that age categories 40-49 y (OR: 7.9, 95% CL: 3.8-16.2) and 50-59 y (OR: 8.0, 95% CL: 3.5-18.2); non-Hispanic blacks (OR: 2.4, 95% CL: 1.3-4.1); less than high school education (OR: 2.6, 95% CL: 1.5-4.8); < 2.0 times the poverty level (OR: 3.5, 95% CL: 1.9- 6.6); heroin consumers (OR: 2.3, 95% CL: 1.1-4.6); IDU (OR: 8.1, 95% CL: 3.1-21); blood transfusion recipients (OR: 2.9, 95% CL: 1.4-5.7) and >/= 10 lifetime sex-partners (OR: 5.5, 95% CL: 1.5-19.7) were significantly associated with CHI.

Multivariate logistic regression analysis in the chronic HCV risk factor model indicated that persons in moderate (OR: 2.5, 95% CL: 1.1-5.5) and high (OR: 30.3, 95% CL: 12.1-76) substance abuse risk factor categories were significantly associated with CHI. Multivariate logistic regression analysis in chronic HCV BRFSS model indicated that age categories 40-49 y (OR: 7.5, 95% CL: 3.5-15.9) 50-59 y (OR: 8.7, 95% CL: 4.2-18); males (OR: 3.1, 95% CL: 1.5-6.4); non-Hispanic black (OR: 1.8, 95% CL: 1.1-2.9); less than high school education (OR: 2.2, 95%CL: 1.3-3.8); < 2.0 times the poverty level (OR: 3.7, 95% CL: 2.0-6.8); alcohol consumers (OR: 1.7; 95% CL: 1.0-2.9) and smokers (OR: 3.7, 95% CL: 1.8-7.6) were significantly associated with CHI. c-statistic of the three models were 0.94, 0.92 and 0.88 respectively thereby implying that all three models were strong models with a higher predictive accuracy of CHI.

Conclusions:

We conclude that the estimated prevalence of CHI in this analysis sample is 0.7%, however the true prevalence estimates of CHI are likely to be significantly higher if incarcerated, homeless and other population not presented in NHANES are included. IDU continues to be the strongest risk factor for CHI. Persons with two or more addictive behavioral risk factors have significant associations with CHI. Results from this study will enable identification of at-risk population for CHI and provide valuable information for linking them to appropriate treatment and care.

CHAPTER ONE

INTRODUCTION

1.1 Background

Hepatitis C virus (HCV) infection is the most common blood-borne infection worldwide. In 1989, Choo et al (1) successfully cloned a single cDNA clone derived from a new flavi-like virus by various molecular biological methods. This virus was responsible for most post-transfusion hepatitis, also called type C hepatitis, parenterally transmitted non-A non-B hepatitis (PT-NANB), non-B transfusion-associated hepatitis, post-transfusion non-A non-B hepatitis, and this virus was identified as HCV (2). Globally, up to 3% of the world's population, which is about 200 million individuals, are estimated to have HCV infection (3); ~71 million are estimated to have CHI out of which 400,000 deaths occur due to cirrhosis and hepatocellular carcinoma because of CHI (World Health Organization Fact Sheet, October 2017). In the civilian non-institutionalized U.S. population, approximately 2.7 million persons have CHI (4).

HCV infection remains asymptomatic in many persons and approximately half the infected are unaware of their infection. Within 30 years, 41% of infected persons' progress to cirrhosis, hepatocarcinoma and mortality from liver-related causes (5). A majority of the infected do not receive antiviral treatment because they are unaware of their infection (6). Due to these reasons, the quality of life (QOL) is negatively affected in the infected population along with exorbitant health care costs associated with treatment and care of the infected. Therefore, it is important to identify the at-risk population for HCV infection and direct them to appropriate treatment and care to positively impact the QOL and decrease the healthcare associated costs. Anti-HCV positive persons are positive to HCV antibody which indicate prior or current Hepatitis C virus infection termed as acutely infected population and HCV RNA positive persons indicate current infection termed as chronically infected population. Chronic HCV infection (CHI) is developed from acute HCV infection in the affected population.

1.2 Purpose of this study

Denniston *et al* (4) have conducted a comprehensive study for determination of risk factors associated with CHI in the US non-institutionalized population between 2003-2010. This is the most comprehensive study conducted to date where associations between behavioral risk

factors and CHI were determined. It would be useful to examine all possible risk factors associated with CHI to better identify the at-risk populations. It would also be useful to better characterize the risk associated with CHI by developing summary measures of risk which could be used in clinic settings and/or in future studies for identification of at-risk individuals for CHI who should be tested and treated. In addition, it will be helpful to estimate CHI prevalence at the state or county levels by using predictive models. In this study, we propose to fill these gaps by conducting a multivariate examination of all possible risk factors of CHI in the non-institutionalized U.S. adult population using National Health and Nutrition Examination Survey (NHANES) datasets from 2003-2014.

1.3 Research Questions

Research questions asked in this study are below,

1. Is there an association between age, sex, race/ethnicity, marital status, education and family income with CHI after adjusting for confounders in the US adult population 20 – 59 years of age?

2. Is there an association between smoking, alcohol consumption, illegal drug use, injection drug use, blood transfusion and lifetime sex-partners with CHI after adjusting for confounders in the US adult population 20 – 59 years of age?

3. Does the odds of CHI increase with increasing number of substance abuse risk factors in the US adult population 20 – 59 years of age?

4. Can the risk for CHI in the US adult population 20 – 59 years of age be modeled without sensitive behavioral information regarding risk factors such as illegal drug use or number of sexual partners so that predictive models of CHI prevalence could be developed from other data sources that do not include such sensitive behavioral information?

1.4 Hypotheses

Our hypotheses are,

1. One or more of the demographical variables in 1.3.1 is significantly associated with CHI by adjusted multivariate analysis.

2. One or more of the behavioral risk factors and/or blood transfusion in 1.3.2 is significantly associated with CHI by adjusted multivariate analysis.

3. Prevalence and odds of CHI significantly increases with associations with increasing number of substance abuse risk factors.

1.5 Organization of the thesis

This thesis is organized into five chapters. Chapter one is the Introduction which describes the background of the study, purpose of this study, research questions and hypotheses. Chapter two is the Literature Review followed by detailed description of the Methodology in Chapter three. Chapter four details the Results and Chapter Five is composed of Discussion, Study Limitations and Conclusions.

CHAPTER TWO

LITERATURE REVIEW

2.1 Hepatitis C is a liver disease caused by the Hepatitis C virus: biology and pathology

HCV first characterized in 1989 by Choo *et al* (1) and Kuo *et al* (7) is an enveloped RNA virus of the genus Hepacivirus of the family Flaviviridae. Its genome consists of 9.6-kb single-stranded RNA which codes for a long polyprotein of approximately 3000 amino acids which is processed post-translationally to yield structural proteins (core and envelope proteins E1, E2) and non-structural (NS) proteins. The envelope proteins are the outer surface proteins of the viral particles which play a key role in virus entry into the host cell (8). HCV RNA virus has a high degree of heterogeneity resulting in six major genotypes and more than 120 subtypes of HCV (9).

There are three major types of HCV genotypes- genotype 1 (GT1), genotypes 2 and 3 (GT2 and GT3) that influence disease progression and responses to therapy. In the US, GT1 is the most prevalent as it is affects ~70% of patients while GT2 and GT3 affect ~13% of patients with HCV.

2.2 Acute and Chronic HCV infection, its comorbidities

Hepatitis C is a contagious liver disease that ranges in severity from a mild illness lasting a few weeks to a serious, lifelong illness that attacks the liver. **Acute Hepatitis C virus infection** is a short-term illness that occurs within the first 6 months after infection with HCV. Acute HCV is usually asymptomatic and about 15-45% of the infected spontaneously clear the virus by a strong immune response without any treatment within 6 months of infection. Acute HCV is a contagious viral infection spread through contact with infected blood and bodily fluids. **Chronic Hepatitis C virus infection** is a long-term illness that occurs when the HCV RNA remains in a person's body for at least 6 months after viral transmission (10). In 55-85% of the population that don't clear the virus after an acute HCV infection, it will develop into a chronic HCV infection (CHI). The presence of hypervariable regions in the E2 envelope glycoprotein, lack of proof reading ability, high rate of generating new viral variants during infection, ability to evade the host immune responses together allow HCV to persist in the infected persons and establish

into a CHI (11). In those people who develop CHI, the infection is often undiagnosed because it remains asymptomatic until decades after infection when symptoms develop and lead to serious liver problems, including hepatitis, cirrhosis (scarring of the liver) or liver cancer (Figure 1). Persistent HCV infection is accompanied by liver cirrhosis, hepatocellular carcinoma (HCC), end stage liver disease and finally death (12).

Lu *etal* (13) carried out The Chronic Hepatitis Cohort Study (CHeCS) that comprised of 11,167 adults with CHI receiving care at one of four large health systems and studied changing trends in rates of cirrhosis, decompensated cirrhosis and all-cause mortality. Results from this study showed that prevalence of cirrhosis increased from 20.8% to 27.6% from 2006 to 2015 in chronic HCV patients. Their study showed that HCV patients >/60 years of age had the highest prevalence of liver-related complications when compared to younger patients <60 years of age. Similar findings were published by El- Serag *et al* (14) from their retrospective cohort study of 161,744 chronic HCV patients in the Veterans Health Administration Hepatitis C Clinical Case Registry. Persistent HCV infection is accompanied by liver cirrhosis, hepatocellular carcinoma (HCC), end stage liver disease and finally death (12). Kanwal *etal* (15) conducted a retrospective cohort study of approximately 110,000 US veterans with CHI of known HCV genotypes 1, 2, 3, and 4 from the VA HCV Clinical Case Registry between 2000 and 2009. Results from this study showed that HCV GT3 was associated with a significantly increased risk of developing cirrhosis and hepatocarcinoma when compared to genotype 1.

Figure 1: Natural history of hepatitis C virus infection (16)

*in patients with liver cirrhosis
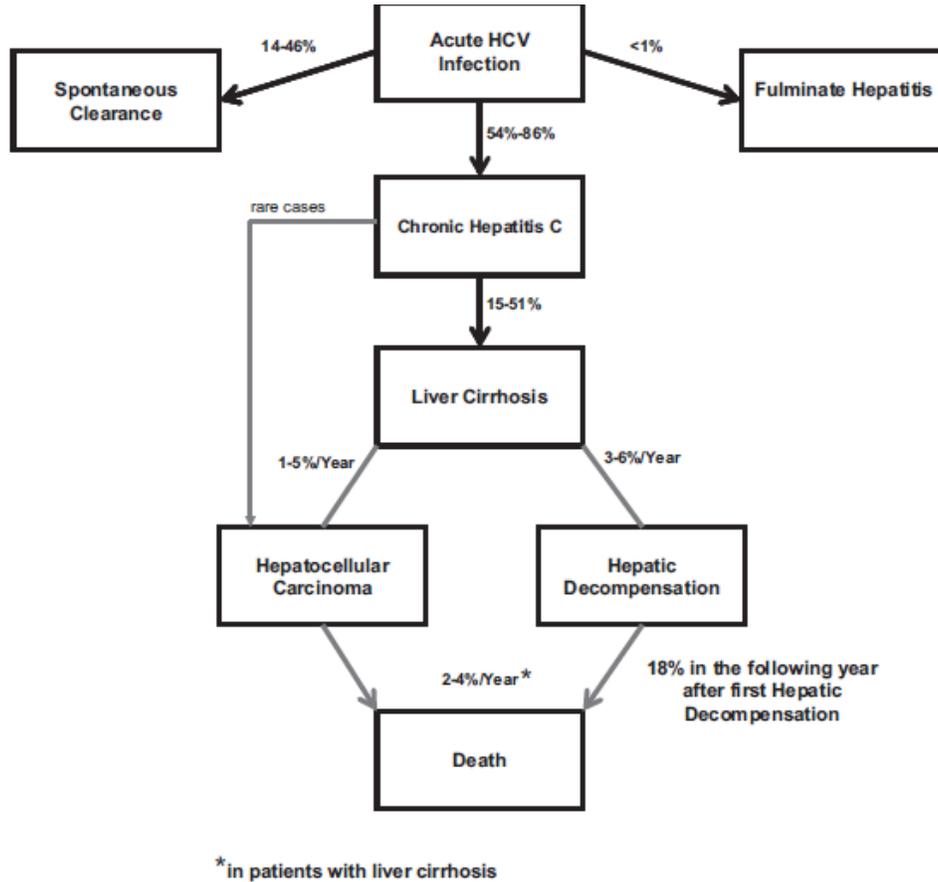
2.3 HCV related healthcare costs

The quality-adjusted life year (QALY) is a generic measure of disease burden, which includes both measures of morbidity (self-reported health) and mortality to assess the quality and quantity of life, lived. It is used in economic evaluation to assess the value for money of medical interventions. One QALY is equal to one year in perfect health. In their study, Younossi *et al* compared the cost-effectiveness (CE) of three HCV screening strategies for treatment with oral direct acting antiviral drugs (DAA). The three strategies were screen all (SA), screen Birth Cohort (BCS), and screen high risks (HRS). SA cost $272.0 billion with 12.19 QALYs per patient, BCS cost $274.5 billion and led to 11.65 QALYs per patient and HRS cost $284.5 billion with 11.25 QALYs per patient. This study concluded that screening the entire US population and treating active viraemia as cost-saving (17).

Davis *et al (18)* conducted a case-control study where they analyzed a large United States claims database (from 1/1/2002 to 12/31/2006) to estimate all-cause and disease-related resource utilization and costs among managed care enrollees with chronic HCV. Use and costs of medical services and prescription drugs over a 12-month post-diagnosis period were determined. Results from this study showed that adjusted all-cause costs were $20,961 per HCV patient when compared with $5451 per control (p<0.0001). Hospitalization occurred in 24% of HCV patients when compared with 7% of controls (p<0.0001). This study concluded that disease-related costs in HCV exceeded all-cause costs in demographically matched controls. Total healthcare cost associated with HCV infection in 2011 was $6.5 ($4.3-$8.2) billion which is expected to peak in 2024 at $9.1 ($6.4-$13.3) billion (19). Majority of this peak cost will be attributable to advanced liver diseases such as decompensated cirrhosis (46%), compensated cirrhosis (20%) and HCC. Hunter *et al* (20) performed a retrospective study of chronic HCV patients to determine the relationships between HCV genotypes and liver disease progression, healthcare resource utilization (HCRU) and healthcare costs. This study showed that patients with GT1 had the highest total all-cause costs, patients with GT3 had the highest liver-related comorbidities and patients with GT2 had lower HCRU and the lowest costs. GT3 is associated with higher risk of liver complications that leads to increased health care cost (21, 22).

In a recent study El Khoury *et al* (23), estimated the burden of untreated HCV infection in patients associated with high economic costs and reduced quality of life in comparison with matched controls. Annual productivity losses of untreated HCV infection patients were significantly higher when compared with matched controls ($8,209 vs. $4,424, p < 0.001) and the average total costs were approximately $27,000 per untreated HCV infected patient per year which was about 150% of the total costs of the matched controls. This study also showed that untreated HCV infected patients had significantly lower health related quality of life (HRQoL) than matched controls.

The advent of DAAs has been revolutionary in the advancement of HCV treatment. DAAs have fewer side effects, shorter duration times (~12 weeks) for treatment, high sustained virologic response (SVR) and are effective regardless of race and gender. Exorbitant prices of DAAs serve as a major obstacle to the wide use of these drugs for HCV treatment. Average cost

per pill is ~$1000 and approximate cost of treatment for 12 weeks is $84,000 (24). However, in August 2017 Food and Drug Administration (FDA) has approved two new DAAs to treat all genotypes of HCV. AbbVie's Mavyret (glecaprevir/pibrentasvir) was approved and launched for a reduced price of $26,400 for 8 weeks (www.mavyret.com). This lower cost is partly due to its shorter treatment duration, but even people who need 12 or 16 weeks would still pay less than they would for existing therapies.

2.4 Estimation of prevalence gives us key information about the scope of the problem

HCV prevalence is highest in Africa and the Middle East ranging from 2 – 15%, whereas prevalence in North America, Japan, Australia, Northern and Western Europe is the lowest not >2%. Egypt has the highest HCV prevalence (25, 26). In the United States non-institutionalized civilian population, NHANES estimated that approximately 2.7 million persons are chronically infected with HCV and about 3.6 million people are acutely infected with HCV between 2003 and 2010 (4). Studies conducted earlier between 1988 and 1994 yielded a similar estimate of the US population with CHI of 2.7 million persons (27) and 3.2 million persons had CHI between 1999 and 2002 (28). However, these numbers underestimate the true prevalence because NHANES does not include institutionalized population, which is at a high-risk of HCV infection. Edlin *et al (29)* conducted a systematic review of peer-reviewed literature and unpublished presentations to estimate the prevalence of hepatitis C in the excluded populations in the US. An estimated 1.0 million HCV antibody positive persons are excluded from the NHANES sampling frame and 0.8 million are currently infected leading to a total of at least 4.6 million with HCV antibody and 3.5 million currently infected.

Overall prevalence of cirrhosis in HCV-infected patients increased from 20.8% in 2006 to 27.6% in 2015 and overall prevalence of decompensated cirrhosis increased from 9.3% in 2006 to 10.4% in 2015 in HCV infected patients aged 60 or older (13). The high prevalence of HCV and HCV mediated comorbidities in the US indicate the persistence of HCV infection since its discovery in 1989 due to numerous reasons inclusive but not limited to higher rates of treatment costs, undiagnosed and/or untreated infections and behavioral risk factors among others.

2.5 Risk Factors for HCV infection – Demographic and Behavioral risk factors

HCV infection in patients is primarily spread by blood contact. Intravenous drug injection is the primary risk factor associated with new HCV infections. Therefore the three primary risk groups are intravenous drug users, recipients of blood transfusion before 1992 and health care workers (30). Since 1992, blood donors are routinely screened for HCV to eliminate HCV infected blood in the process of blood transfusion. Alter *et al* (27) were the first to study prevalence of HCV in the US using NHANES datasets by behavioral risk factors. In this study conducted to estimate prevalence of HCV from 1988 – 1994, the authors determined that illegal drug use and high-risk sexual behavior were the strongest risk factors independently associated with HCV infection among participants 17 to 59 years of age. Increased prevalence of HCV infection was associated with increasing number of times cocaine or marijuana was used and highest prevalence of HCV infection was seen among persons who had 10 or more sexual partners.

HCV seroprevalence has been reported in 75-90% of long-term (>3 years) injection drug users (IDU) and in 18%-38% of short-term (<3 years) IDUs (31-33). It is also a known fact that injection drug use is the strongest risk factor independently associated with HCV infection. For example, the 1945-1965 birth cohort is an important predictor for anti-HCV positivity because of high rates of injection drug use in this cohort termed the baby boomers' cohort. In 2012, CDC issued a recommendation to test all persons born in this cohort for HCV infection without prior risk ascertainment (34). This was further supported by the findings from Smith *et al* (35) who determined the prevalence and predictors of anti-HCV positivity among primary care outpatients using risk-based testing that 74% of the identified anti-HCV positive patients were born between 1945-1965. Anti-HCV positivity was significantly higher in these patients when compared with the referent group of those born before 1945 or after 1965.

Patients undergoing hemodialysis are at a higher risk for HCV infection with a prevalence estimate of 0.8% of HCV antibodies in the US (36). Prevalence in hemodialysis patients has been found to increase with longer hemodialysis duration, male sex, black ethnicity, comorbidities (example: diabetes, hepatitis B), prior kidney transplant, alcohol or drug abuse (37). Demographic risk factors such as male sex, older age of 40-59 years, non-hispanic black, lesser

than high school education and lower family income were determined to be significantly and independently associated with HCV infection by Denniston *et al* (4).

2.6 Public health impact of examination of comprehensive behavioral risk factors for CHI

Various studies listed in this review have examined one or more of the risk factors associated with HCV infection inclusive of demographic factors such as age 40-59 years, male sex, lower family income, non-Hispanic black ethnicity, 1945-1965 birth cohort and behavioral factors such as illegal drug use, alcohol use and >/= 10 life time sex-partners. Denniston *et al* (4) conducted a comprehensive study by examination of the aforementioned demographic risk factors, illicit drug use and receipt of blood transfusion before 1992 and determined that persons aged 40-59 years, non-Hispanic black, less than high school education, illicit drug use and receipt of blood transfusion before 1992 were significantly associated with CHI.

Comprehensive risk factor profile may include a combination of different types of illicit drugs, injection drugs, alcohol use, smoking, and >10 lifetime sex-partners. Examination of comprehensive profile of behavioral risk factors for CHI will help to refine identification of the populations at risk for CHI, enabling screening, behavioral interventions, and linking them to care and treatment. Up-to-date, an estimation of CHI prevalence using comprehensive behavioral risk factor profile has not been published. In our study, we propose to estimate the prevalence of CHI in the US non-institutionalized, civilian population using NHANES datasets from 2003-2014. We will determine comprehensive behavioral risk factors associated with CHI, which will improve identification of at risk populations to link them to optimal treatment, care, prevention of comorbidities and reduce the associated health care costs. Estimates of CHI burden are essential to guide policy and programs to optimally prevent, detect and treat the infection. State-level estimates of the CHI prevalence are essential for developing intervention programs, research, and federal assistance funding priorities among US states. Results from this thesis will be critical to develop a prediction model for estimation of state-level prevalence of CHI where less information on behavioral risk factors are available.

CHAPTER THREE

METHODOLOGY

3.1 Data Source

We performed secondary data analyses using the Centers for Disease Control and Prevention's National Center for Health Statistics survey called NHANES. This survey collects nationally representative data on the health and nutritional status of the U.S. noninstitutionalized civilian population and is a public use dataset. In this study, we combined data from several NHANES data sources including demographic information, laboratory testing and questionnaire datasets from subjects interviewed between 2003-2014. NHANES provides information on the non-institutionalized civilian resident population. It excludes persons in care or custody of institutionalized settings, all active-duty military, all active- duty family members living overseas, and US citizens living outside of the 50 states and District of Columbia (38).

3.2 Survey and Sample design

NHANES examines a nationally representative sample of about 5,000 persons each year located in counties across the US. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel. Findings from the survey was used to determine the prevalence of major diseases and risk factors for diseases.

A four-stage sample design was implemented in the datasets used in this study. The first stage primary sampling units (PSUs) were selected from a frame of all U.S. counties; the second-stage included a sample of area segments such as census blocks; the third-stage sample selection consisted of dwelling units (DUs) such as noninstitutionalized group quarters and the fourth stage consisted of persons within DUs or households (38).

3.3 Demographics Variables

Sample persons demographics data files for 2003-2004, 2005-2006, 2007-2008, 2009-2010, 2011-2012 and 2013-2014 were downloaded from the NHANES website for each cohort

(https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/DEMO_C.htm). Following is the list of demographic variables used in this study,

1. RIDAGEYR- Age and screening Adjudicated, categorical variable. This is the age of the sample person at the time of the screening interview. From 1 through 84 years of age, age of the sample person is reported by single year of age. Adults 85 years and older have a value of 85. In this study, age categories were 0-19, 20-39, 40-59, 60 plus.

2. RIDRETH1- Race/Ethnicity. This categorical variable was derived from responses to the survey questions on race and Hispanic origin. In this study, all other races except non-Hispanic black were recoded into a 'Other Races' category.

3. RIAGENDR- Gender, categorical variable. Females were the referent group in adjusted and unadjusted logistic regression models in this study.

4. DMDEDUC2- Education level, adults 20+, continuous variable. This variable is the highest grade or level of education completed by adults 20 years and older. Range of value descriptions are less than 9th grade, 9-11th grade, high school/ General Equivalency Diploma (GED), some college or associate degree, college graduate or above. Response categories were recoded into less than high school graduate/GED education and greater than high school graduate/GED education.

5. DMDMARTL- Marital Status, categorical variable. Response categories in this variable were recoded into i). Never married, ii). Widowed/Separated/Divorced and iii). Married/Living with partner.

6. INDFMPIR- Family Poverty to Income Ratio, categorical variable. This variable is an index for the ratio of family income to poverty. This variable was calculated by dividing family income by the poverty guidelines specific to the family size, state and year. Response values in this variable were recoded into i). < 2.0 times the family income to poverty ratio and ii). >/= 2.0 times the family income to poverty ratio.

3.4 Behavioral Risk Factor Variables

Sample persons questionnaire data files for behavioral risk variables were downloaded from respective cohorts from NHANES website (https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&Cycle

). Alcohol use, drug use, smoking and sexual behavior variables were the behavioral risk variables while blood transfusion was the medical examination variable in this study.

Blood Transfusion variable

I.   MCQ092: Ever receive blood transfusion

Alcohol variables

I.   Alq101: Had at least 12 alcohol drinks/year

II.   Alq110: Had at least 12 alcohol drinks/lifetime

III.   Alq130: Average number of alcoholic drinks/day in the past 12 months

Four alcohol categories were created by recoding the alcohol variables as follows (39),

- Alcohol category 1: Lifetime abstainers < 12 drinks ever
- Alcohol category 2: Former drinkers >/= 12 drinks in their lifetime but none in the past year
- Alcohol category 3: Non-excessive current drinkers on average reported
  - Male gender: </= 14 drinks/week
  - Female gender: </= 7 drinks/week
- Alcohol category 4 = Excessive current drinkers reported
  - Male gender: > 14 drinks/week
  - Female gender: > 7 drinks/week

These were further condensed into two alcohol risk categories where alcohol categories 1, 2, 3 were classified as sample persons with no alcohol risk and alcohol category 4 was classified as sample persons with alcohol risk.

Drug variables

For 2003-2004 cohort following variables were recoded into ever-inject variable in this study,

I.   Duq120: Ever used needle to take drugs

II.   Duq100: Ever used cocaine or other street drug

Variables for use of marijuana or hashish, cocaine, heroin and methamphetamine were not available for this cohort.

For 2005-2014 cohorts following variables were used which were recoded into variables as indicated below,

I. Duq200: Ever used marijuana or hashish which was recoded as evermarijuana

II. Duq240: Ever used cocaine/heroin/methamphetamine

III. Duq250: Ever used any form of cocaine along with duq240 were recoded as evercocaine

IV. Duq290: Ever used heroin along with duq240 were recoded as heroin

V. Duq330: Ever used methamphetamine along with duq240 were recoded as evermethamphetamine

VI. Duq370: Ever used needle to inject illegal drug was recoded as everinject

Smoking variable

I. Smq020: Smoked at least 100 cigarettes in life was recoded as ever_smoked

Sexual behavior

For all cohorts, following variables were recoded into lifetime sex-partners (Table 1).

Table 1: Lifetime sex-partners variable and recode

Cohort	Variable name for number of female sex-partners/lifetime (A)	Variable name for number of male sex-partners/lifetime (B)	Variable recode into lifetime sex-partners
2003-2004	Sxq170	Sxq200	A+B
2005-2014	Sxq171	Sxq101	A+B

3.5 Laboratory Testing

Laboratory testing was carried out to detect anti-HCV in blood or serum using direct solid-phase enzyme immunoassay (VITROS Anti-HCV Immunodiagnostic System, Ortho Clinical Diagnostics, Rochester, New York). A confirmatory recombinant immunoblot assay (RIBA) (RIBA

HCV 3.0 Strip Immunoblot Assay, Emeryville, California) an in vitro qualitative immunoassay was performed for the detection of anti-HCV in human blood or serum. Samples with positive results on RIBA testing were confirmed as positive for anti-HCV, with negative results were reported as confirmed negative for anti-HCV and those with indeterminate results were reported as indeterminate.

Chronically infected persons are currently infected and it is important in clinical practice to identify these persons. Serum samples that were confirmed positive or indeterminate for anti-HCV were further tested for HCV RNA using COBAS AMPLICOR HCV Test, version 2.0 (Roche Diagnostics, Indiana, US) an *in-vitro* nucleic acid amplification for HCV RNA on the COBAS AMPLICOR Analyzer (Roche Diagnostics) for samples from 2003 to 2010, 2012-2014 and the AMPLIPREP COBAS TaqMan HCV Test performed on the COBAS AMPLIPREP and COBAS TaqMAN 48 Analyzer for samples from 2011-2012 (4).

Sample persons laboratory data files for HCV RNA were downloaded from respective cohorts from the NHANES website (https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory&CycleBeginYear=2003). For 2003-2004, SSHCVRNA was the HCV RNA variable. Exam sample weights were used for the analysis of this cohort. For rest of the cohorts LBXHCR was the HCV RNA variable used without the exam sample weights since it was not required.

I. SSHCVRNA: Hepatitis C RNA
II. LBXHCR: Hepatitis C RNA

3.6 Statistical Analyses:

SAS version 9.4 developed by the SAS Institute (NC, USA) a statistical package designed to analyze complex survey data was used for all analyses in this study. Estimates were weighted to represent the total U.S. noninstitutionalized civilian population, to account for oversampling and nonresponse to the household interview and medical examination. Two-year sample weights (WTMEC2R) were used for the weighted analyses. A *p* value less than 0.05 was considered statistically significant.

For descriptive statistics, to estimate the prevalence of CHI persons with demographic characteristics such as age such as smoking, alcohol use, illegal drugs use (marijuana, cocaine, methamphetamine, heroin and injection drugs), blood transfusion and number of lifetime sex-partners we have implemented bivariate analyses. We created an ordinal variable with the number of risk factors - 0, 1, 2, 3, 4, 5, 6, 7 for calculation of risk factor scores for CHI.

We have conducted unadjusted and adjusted weighted multivariate logistic regression analyses for determination of risk factors significantly associated with CHI in persons of age 20 to 59 years because data on drug use and sexual behaviors in persons < 20 years and >/= 60 years of age are not available from NHANES.

We have included and analyzed missing data cases as a separate domain rather than excluding the data from the analyses because missing data is not always completely at random and its inclusion accounts for the accurate analyses of the entire data sets. We have used the Not Missing Completely at Random (NOMCAR) option in weighted logistic regression models to include the missing data. The following groups were used as referent groups for the analysis variables used in logistic regression models:

1. Age = 20-39 y
2. Sex = Female
3. Race/Ethnicity = All others except non-Hispanic black
4. Marital = Married, living with partner
5. Education= High school or more
6. Family income = >= 2.0 times poverty level
7. Smoking = Never smoked
8. Alcohol consumption = combined 3 groups (i.e., never drinkers, former drinkers, and non-excessive current drinkers)
9. Marijuana = Never
10. Cocaine = Never
11. Heroin = Never
12. Methamphetamine = Never
13. Injection drug use (IDU) = Never

14. Blood Transfusion = No

15. Lifetime Sex-partners = 0-1 lifetime sex-partners

For unadjusted and adjusted multivariate logistic regression analyses, we have pursued three models to determine the association between risk factors and CHI. They are

1). **Chronic HCV full model (Model 1) -** In this model, we have included all the demographic and behavioral risk factors mentioned in this study to assess the associations with CHI.

2**). Chronic HCV risk factor model (Model 2)-** In this model, we have included the demographic variables, blood transfusion, number of CHI.

3). **Chronic HCV Behavioral Risk Factor and Surveillance System (BRFSS) model (Model 3)-** BRFSS is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions and use of preventive services (www.cdc.gov/brfss/index.html). In this model, we have included only variables that are also available on the BRFSS. This includes the demographic variables, alcohol use and smoking to assess the associations with CHI. In the future, the results of model will be critical for developing a prediction model to estimate state-level prevalence of CHI.

The C-statistic (sometimes called the 'concordance' statistic or C-index) is a measure of goodness of fit for binary outcomes in a logistic regression model. It is a standard measure of the predictive accuracy of a logistic regression model (40). It is equal to the area under the Receiver Operating Characteristic (ROC) curve and ranges from 0.5 to 1.

- A value < 0.5 indicates a very poor model.
- A value of 0.5 means that the model is no better than predicting an outcome by random chance.
- Values > 0.7 indicate a good model.
- Values > 0.8 indicate a strong model.
- A value of 1 means that the model perfectly predicts those group members who will experience a certain outcome and those who will not.

We have compared the c-statistics between the three models in this study.

CHAPTER FOUR

RESULTS

4.1 Estimated Prevalence of HCV RNA

The estimated number of HCV RNA-positive persons among of age 20-59 years in birth cohorts 2003-2004, 2005-2006, 2007-2008, 2009-2020, 2011-2012, 2013-2014 were 2.56 million (95% CI, 1.75-3.36), 2.19 million (95% CI, 1.23-3.13), 2.68 million (95% CI, 1.71-3.64), 1.95 million (95% CI, 1.12-2.76), 2.08 million (95% CI, 0.95-3.19) and 1.92 million (95% CI, 1.41-2.43) respectively (Figure 2). Figure 2 shows that estimated number of CHI population has decreased since 2008, but the differences were not statistically significant.

Figure 2: Weighted estimates of CHI persons of age 20-59 y from 2003-2014
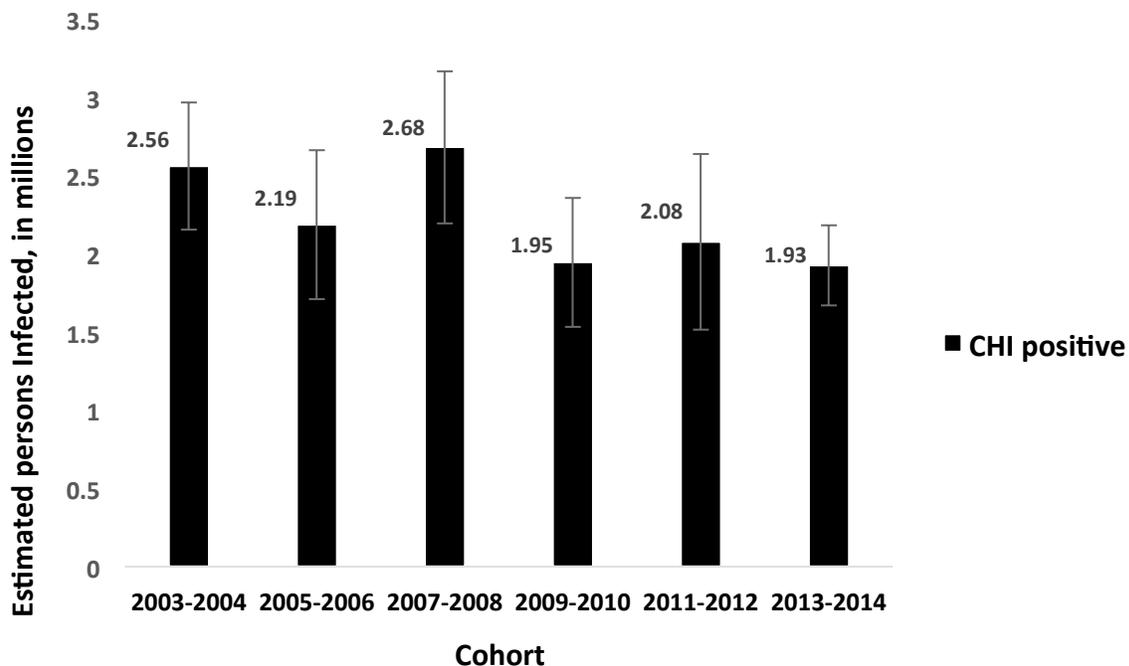
Drug use, smoking and alcohol use were the substance abuse risk factors for CHI in this study. Based on the number of substance abuse risk factors, we grouped these into three categories as low (0-1), moderate (2-4) and high (5-7). Next, we estimated the prevalence of chronic HCV RNA-positive persons by these three categories (Figure 3). Prevalence of CHI increased

gradually from 0.2% at low to 1.1% at moderate risk, followed by a sharp increase to 11.0 % at high-risk population.

Figure 3: Prevalence of chronic HCV by number of substance abuse risk factors, 2003-2014
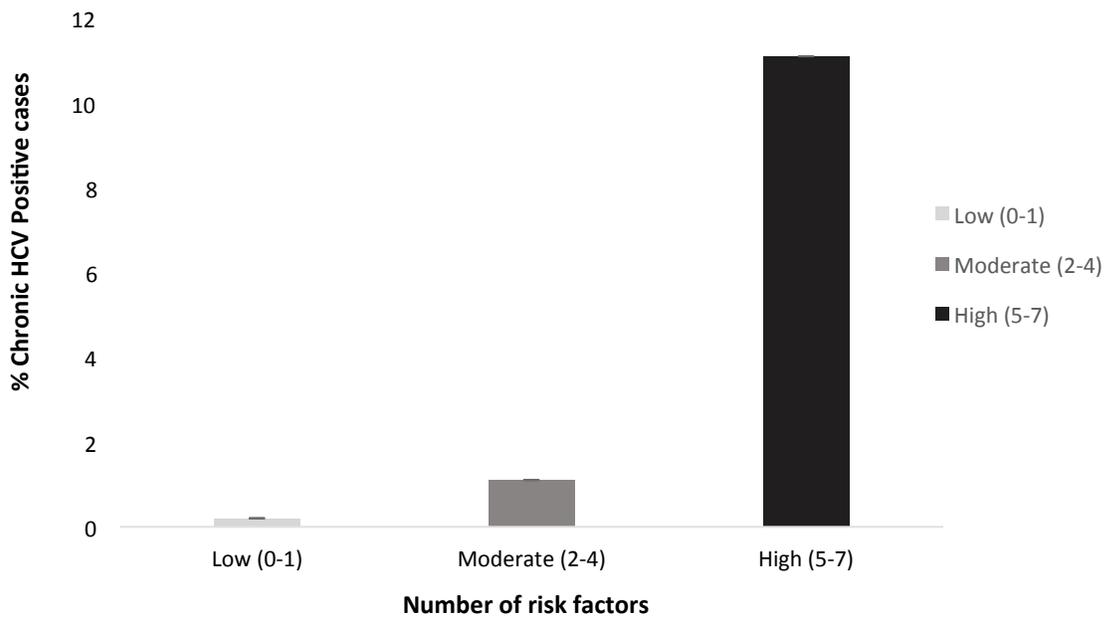
4.2 Crude associations of Demographic and Behavioral Risk Factors of persons with CHI

We have presented descriptive statistics on demographical and behavioral risk factors of persons with CHI by bivariate analyses (Table 2). Age, race/ethnicity, education, marital status, and family income were associated with HCV-RNA-positive status.

Participants of age 40-59 years had a CHI prevalence of 2.48% and more likely to be infected, whereas CHI prevalence was 0.37 in 20-39 y age group. Non-Hispanic blacks had a CHI prevalence of 2.25% when compared to 1.14% in other races. Male participants had a CHI prevalence of 1.84% when compared to 0.96% in females. Participants with less than high school education/GED had a CHI prevalence of 2.26% when compared to 1.11% in participants who had high school education and above. Widowed/divorced/separated participants had the highest CHI prevalence of 2.75% when compared to never married, married/living with partner groups who had a CHI prevalence of 1.00% and 1.20% respectively. Participants who were at < 2.0 times poverty level had a CHI prevalence of 2.15% when compared to 0.76% in participants who were at > 2.0 times poverty level.

Table 2: Demographic Characteristics by CHI status in participants of age 20-59 years, NHANES 2003-2014

Variable	Category	Sample Size	HCV RNA-Positive	%HCV RNA-positive
Age, years	20-39	10,696	40	0.37
	40-49	5,309	111	2.09
	50-59	4,649	134	2.88
Race	Non-Hispanic black	4397	99	2.25
	Other races	16,257	186	1.14
Gender	Male	9,907	182	1.84

	Female	10,747	103	0.96
Education	Less than high school education	4,787	108	2.26
	High school education and above	15,851	176	1.11
Marital status	Never married	5,007	50	1.00
	Widowed, divorced, separated	3,052	84	2.75
	Married, living with partner	12,585	151	1.20
Family Income	< 2.0 times poverty level	9,158	197	2.15
	>/= 2.0 times poverty level	10,044	76	0.76

Bivariate analysis of behavioral risk factors indicated that participants who smoked had a CHI prevalence of 2.76% as compared to 0.33% among non-smokers. Participants who consumed alcohol had a CHI prevalence of 1.71% and more likely to be infected when compared to non-consumers of alcohol who had a 0.84% CHI prevalence. Participants who had consumed marijuana, cocaine, methamphetamine, heroin had a CHI prevalence of 2.23%, 5.01%, 7.17% and 18.16% when compared to the CHI prevalence 0.4%, 0.55%, 0.93% and 0.94% among non-consumers of marijuana, cocaine, methamphetamine and heroin respectively. IDU participants had a CHI prevalence of 30.24% while non-injection drug users were 0.71% CHI prevalent. Recipients of blood transfusion had a CHI prevalence of 2.47% as compared to 1.19% in participants who were not recipients of blood transfusion. Participants who had >/= 10 lifetime sex-partners had a CHI prevalence of 2.63% as compared to 0.27% and 0.82% in other

groups. Participants in high substance abuse risk factors category had a CHI prevalence of 11.11% as compared to 0.24% and 1.12% CHI prevalence in participants in low and moderate categories (Table 3).

Table 3: Behavioral Risk Factors by CHI status in participants of age 20-59 years, NHANES 2003-2014

Variable	Category	Sample Size	HCV RNA-Positive	%HCV RNA-positive
Ever smoked	Yes	8,911	246	2.76
	No	11,734	39	0.33
Alcohol	Yes	7,708	132	1.71
	No	9,239	78	0.84
Evermarijuana	Yes	8,424	188	2.23
	No	7,079	28	0.40
Evercocaine	Yes	2,837	142	5.01
	No	12,651	70	0.55
Evermethamphetamine	Yes	1,130	81	7.17
	No	14,359	133	0.93
Everheroin	Yes	391	71	18.16
	No	15,095	142	0.94
Everinject	Yes	410	124	30.24
	No	17,744	126	0.71
Blood transfusion	Yes	4,170	103	2.47
	No	19,017	227	1.19

Lifetime sex-partners	0-1	2,588	7	0.27
	2 to 9	7,299	60	0.82
	>/=10	5,066	133	2.63
Substance abuse risk factors	Low (0-1)	7,072	17	0.24
	Moderate (2,3,4)	6,352	71	1.12
	High (5,6,7)	702	78	11.11

4.3 Associations between risk factors and CHI adjusted for potential confounders

In this study, we have predicted associations between risk factors and CHI using three models whose results are described below.

1**). Model 1** - For adjusted weighted multivariate logistic regression analysis, demographic and risk factor variables were adjusted for all variables in the model. Simple unadjusted logistic regression indicated for participants of age 20-59 years found that age 40-59 years, male sex, non-Hispanic black, separated/divorced/widowed/living separately, less than high school/GED education, family income less than twice the poverty level, smoking, alcohol consumption, marijuana, cocaine, heroin use, use of injection drugs, blood transfusion and having >/= 10 lifetime sex-partners were significantly associated with CHI.

Adjusted logistic regression analysis for participants of age 20-59 years found that participants with the following characteristics had significant associations with CHI, indicating they had higher odds of an HCV infection when compared to their respective referent groups (Table 4).

1. Age groups 40-49 y and 50-59 y had an odds ratio (OR) of 7.9 (95% CI, 3.8-16.2) and 8.0 (95% CI, 3.5-18.2) respectively
2. Non-Hispanic blacks had an OR of 2.4 (95% CI, 1.3-4.1) as compared to the referent group
3. Less than high school/GED had an OR of 2.6 (95% CI, 1.5-4.8)
4. Family income at < 2.0 times poverty level had an OR of 3.5 (95% CI, 1.9-6.6)
5. Heroin consumers had an OR of 2.3 (95% CI, 1.1-4.6)
6. IDU had an OR of 8.1 (95% CI, 3.1-21)
7. Blood transfusion had an OR of 2.9 (95% CI, 1.4-5.7)
8. >/= 10 lifetime sex-partners had an OR of 5.5 (95% CI, 1.5-19.7)

Table 4: Model 1 - Unadjusted and adjusted odds ratios for the risk associated with CHI in participants of age 20-59 years, NHANES 2003 – 2014 (n = 11,596).

Characteristic	Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI
Age at interview				
20-39 y	(ref)	(ref)	(ref)	(ref)
40-49 y	**5.8**	**2.9-11.5**	**7.9**	**3.8-16.2**
50-59 y	**5.9**	**3.0-11.9**	**8.0**	**3.5-18.2**
Sex				
Male	**2.8**	**1.6-4.9**	2.3	0.9-5.5
Female	(ref)	(ref)	(ref)	(ref)
Race/ethnicity				
Non-Hispanic Black	**2.4**	**1.5-3.8**	**2.4**	**1.3-4.1**
Other races	(ref)	(ref)	(ref)	(ref)
Marital				
Married, living with partner	(ref)	(ref)	(ref)	(ref)
Sep/Div/Wid/Liv Sep	**2.7**	**1.5-4.7**	1	0.4-2.1
Never married	1.2	0.7-2.3	1.2	0.5-2.7
Highest education level				
Less than high school/GED	**4.6**	**3.0-7.1**	**2.6**	**1.5-4.8**
High school or more	(ref)	(ref)	(ref)	(ref)
Family income				

< 2.0 times poverty level	**4.6**	**2.8-7.8**	**3.5**	**1.9-6.6**
>= 2.0 times poverty level	(ref)	(ref)	(ref)	(ref)
Smoking				
Yes	**6.7**	**3.5-12.9**	1.9	0.8-4.2
No	(ref)	(ref)	(ref)	(ref)
Alcohol Consumption				
Yes	**2.1**	**1.3-3.3**	1.3	0.7-2.3
No	(ref)	(ref)	(ref)	(ref)
Marijuana				
Yes	**3.6**	**2.0-6.6**	0.7	0.3-1.9
No	(ref)	(ref)	(ref)	(ref)
Cocaine				
Yes	**10.5**	**6.3-17.6**	1.8	0.9-3.7
No	(ref)	(ref)	(ref)	(ref)
Heroin				
Yes	**28.7**	**18.0-45.7**	**2.3**	**1.1-4.6**
No	(ref)	(ref)	(ref)	(ref)
Methamphetamine				
Yes	**12**	**7.2-20.0**	1.9	0.8-4.5
No	(ref)	(ref)	(ref)	(ref)
Injection drugs				
Yes	**46**	**29.3-72**	**8.1**	**3.1-21**
No	(ref)	(ref)	(ref)	(ref)

Blood transfusion				
Yes	**2.9**	**1.7-5.0**	**2.9**	**1.4-5.7**
No	(ref)	(ref)	(ref)	(ref)
Lifetime sex-partners				
0-1	(ref)	(ref)	(ref)	(ref)
2-9	**4.4**	**1.5-13**	2.8	0.9-8.5
>/=10	**22.4**	**7.7-65**	**5.5**	**1.5-19.7**

2**). Model 2**- For adjusted weighted multivariate logistic regression analysis, number of substance abuse risk factors were adjusted for cohort, age, race, gender, education, family income, marital status, blood transfusion and number of lifetime sex-partners (Table 5). Simple unadjusted and adjusted logistic regression indicated that having >/= 2 risk factors were significantly associated with CHI. In the adjusted analysis, participants in moderate and high categories had an OR of 2.5 (95% CI, 1.1-5.5 and 30.3 (95% CI, 12.1-76) respectively, indicating a significant association with CHI when compared to the referent group.

Table 5: Model 2 - Unadjusted and adjusted odds ratios for the risk associated with CHI in participants of age 20-59 years, NHANES 2003 – 2014 (n = 11,596).

Characteristic	Unadjusted		Adjusted	
Number of Drugs, Smoking and alcohol	**OR**	**95% CI**	**OR**	**95% CI**
Low (0-1)	(ref)	(ref)	(ref)	(ref)
Moderate (2-4)	**4.5**	**2.1-9.54**	**2.5**	**1.1-5.5**
High (5-7)	**54.5**	**25.2-117.8**	**30.3**	**12.1-76**

3**). Model 3** - For adjusted weighted multivariate logistic regression analysis, demographic and risk factor variables were adjusted for all variables in the model (Table 6). Simple unadjusted logistic regression for participants of age 20-59 years indicated that 20 – 59 years, male sex, non-Hispanic black, separated/divorced/widowed/living separately, having less than high school/GED education, family income less than twice the poverty level, alcohol consumption and smoking were significantly associated with CHI.

In the adjusted analysis, participants with the following characteristics had significant associations with CHI therefore indicating that they had higher odds of an HCV infection when compared to their respective referent groups.

1. Age groups 40-49 y and 50-59 y had an odds ratio (OR) of 7.5 (95% CI, 3.5-15.9) and 8.7 (95% CI, 4.2-18).
2. Males had an OR of 3.1 (95% CI, 1.5-6.4)
3. Non-Hispanic blacks had an OR of 1.8 (95% CI, 1.1-2.9)
4. Less than high school/GED had an OR of 2.2 (95% CI, 1.3-3.8)
5. Family income < 2.0 times the poverty level had an OR of 3.7 (95% CI, 2.0-6.8)
6. Alcohol consumers had an OR of 1.7 (95% CI, 1.0-2.9) and
7. Smokers had an OR of 3.7 (95% CI, 1.8-7.6)

Table 6: Model 3 - Unadjusted and adjusted odds ratios for the risk associated with CHI in participants of age 20-59 years, NHANES 2003 – 2014 (n = 11,596).

Characteristic	Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI
Age at interview				
20-39 y	(ref)	(ref)	(ref)	(ref)
40-49 y	**5.8**	**2.9-11.5**	**7.5**	**3.5-15.9**
50-59 y	**6**	**3.0-11.9**	**8.7**	**4.2-18**
Sex				
Male	**2.8**	**1.7-4.9**	**3.1**	**1.5-6.4**
Female	(ref)	(ref)	(ref)	(ref)
Race/ethnicity				
Non-Hispanic Black	**2.3**	**1.5-3.8**	**1.8**	**1.1-2.9**
Other races	(ref)	(ref)	(ref)	(ref)
Marital				
Married, living with partner	(ref)	(ref)	(ref)	(ref)
Sep/Div/Wid/Liv Sep	**2.7**	**1.5-4.7**	1.4	0.7-2.8
Never married	1.2	0.7-2.3	1.4	0.7-2.9
Highest education level				
Less than high school/GED	**4.6**	**3.0-7.1**	**2.2**	**1.3-3.8**
High school or more	(ref)	(ref)	(ref)	(ref)
Family Income				
<2.0 times poverty level	**4.6**	**2.8-7.8**	**3.7**	**2.0-6.8**
>= 2.0 times poverty level	(ref)	(ref)	(ref)	(ref)

Alcohol Consumption				
Yes	**2.1**	**1.3-3.3**	**1.7**	**1.0-2.9**
No	(ref)	(ref)	(ref)	(ref)
Smoking				
Yes	**6.7**	**3.5-12.9**	**3.7**	**1.8-7.6**
No	(ref)	(ref)	(ref)	(ref)

4.4 C-statistic- C-statistic of the models 1, 2 and 3 were 0.94, 0.92 and 0.88 respectively indicating that all three models were strong models with a higher predictive accuracy of CHI in persons strongly associated with respective demographic and behavioral risk factor variables in the models.

CHAPTER FIVE

DISCUSSION AND CONCLUSIONS

DISCUSSION

The primary purpose of this study was complex examination of the NHANES datasets from 2003-2014 to assess combination of demographical and behavioral risk factors associated with CHI in the general U.S. population. The NHANES survey provides important information as it comprises the nationally representative sample of the US non-institutionalized population. The standardization of its methods allows for consistent and good quality data collection (41). This is the first study of its kind to use a complex data set of 6 cohorts from 2003 through 2014 for examination of all possible risk factors (demographical + behavioral) significantly associated with CHI in the adult US population 20-59 years of age using 3 models.

As of 2014, we estimated 1.93 million CHI persons in the general U.S. population of age 20-59 years with a prevalence of 0.7% sampled by NHANES. This has slightly declined since 2010 based on Denniston *et al* estimation of 2.7 million as CHI infected persons from 2003-2010 in the general U.S. population (4). As indicated in their study, our analysis also suggests that declining prevalence of CHI in the noninstitutionalized U.S. population may likely be because of increasing mortality from HCV-related conditions (42). This prevalence is an underestimation of the true CHI population in the US because NHANES does not include high-risk populations including the incarcerated, hemodialysis patients, the homeless and people living on Indian reservations, all active military and U.S. citizens living outside the U.S.A.

In this study, substance abuse risk factors such as drug use, alcohol consumption and smoking were grouped into low (0-1), moderate (2-4) and high (5-7) summary risk factor score categories based on the number of risk factors. In the absence of categorization of the summary risk factor scores, fewer number of CHI observations were observed in each summary risk factor score that resulted in wider confidence intervals. Thus, categorization helped to overcome this concern. We have estimated the prevalence of CHI in these categories and found a linear increase in CHI prevalence with increasing number of substance abuse risk factors (Figure: 3) with a maximum of 11% CHI prevalence in persons pf the high category. This is the

first study of its kind to estimate the CHI prevalence in persons associated with one or more substance abuse risk factors.

Higher prevalence of CHI was seen in participants with the following socio-demographic characteristics- 40-59 years, non-Hispanic black, male sex, less than high school education, widowed/divorced/separated and family income less than 2.0 times poverty level when compared to their respective control groups (Table 2). One or more of these observations were also reported in earlier studies (4, 28, 41). The Centers for Disease Control and Prevention and US Preventive Services Task Force have recommended 1-time HCV screening in persons born between 1945-1965 (43, 44) which corresponds to age 49-69 in the year 2014 and account for most of the prevalent HCV cases. Table 3 in this study noticeably shows that the age groups 40-49 and 50-59 had a prevalence of 2.09% and 2.88% respectively with a total prevalence of 4.97% contributing to 93% of the infected population in 20-59 years of age. Results from our study supports the findings from earlier studies (4, 41).

Higher prevalence of CHI was seen in participants with the following behavioral risk characteristics- smoking, alcohol consumption, consumption of marijuana, cocaine, methamphetamine, heroin, injection drug users, blood transfusion recipients prior to 1992 and >/=10 life time sex-partners when compared to their respective referent groups (Table 3). Denniston *et al* (4) in their study have determined higher prevalence of CHI in participants of age 60 and older who had received blood transfusion before 1992, in injection drug users, and with 20-49 life time sexual partners when compared to their respective referent groups. Taylor *et al* (39) have indicated higher prevalence of CHI in participants who were former drinkers and excessive current drinkers when compared to non-drinkers. Results from these studies are in line with some of the observations we have seen in our study although our study is unique because we have estimated CHI prevalence in population who are smokers, consumers of heroin, cocaine and methamphetamine that have not been determined in previous studies.

Knowledge of risk factors for CHI is important for several reasons not limited to 1). Identification of at-risk populations to link to treatment and care 2). Resource allocation for prevention measures and 3). To propose policy and guidelines for control and prevention of CHI. This study is the largest and most comprehensive so far where the multivariate models

have included the most established confounders. This study is also the first of its kind to have determined the risk factors associated with CHI by three models. As reported in a previous study (4), the chronic HCV full model (Table 4) elicited that age 40-59 years, non-Hispanic blacks, less than high school/GED, < 2.0 times poverty level were significantly associated with CHI. Additionally, the chronic HCV full model has also established significant associations of heroin consumption, blood transfusion and >/=10 lifetime sex-partners with CHI that have not been reported earlier. IDU remains the strongest risk factor for CHI as reported earlier (4) however in the earlier study IDU was combined with other drugs as a risk factor. For IDU association with CHI, Denniston *et al* (4) have reported an OR of 8.7 whereas we have reported an OR of 8.1, therefore showing that in both the studies the OR in the multivariate model is similar. Race, low socioeconomic status and less education were associated with CHI because these factors are often associated with high-risk behaviors and hence higher risk of infection.

Chronic HCV risk factor model undoubtedly shows that the OR increases with increasing number of risk factors in a person. This is the first study where the number of substance abuse risk factors (alcohol consumption, smoking and drug use) were categorized into low, moderate and high-risk categories to determine the odds of being positive for CHI (Table 4). This data will be crucial in identification of the positively infected population, link them to treatment, and care before they become susceptible to secondary conditions such as hepatocarcinoma and cirrhosis thereby reducing the HCV burden and its associated healthcare costs. Missing values in the druq use questions in NHANES may not be equally distributed between the different socio-economic status (SES) groups, which can lead to the biases in prevalence estimation and/or risk factor estimation.

BRFSS datasets does not contain questions on risk factor variables such as drug use including IDU and number of lifetime sex-partners. We carried out the third and final model to establish a prediction model for the state level estimates of CHI including only alcohol and smoking risk factor variables, which are available in BRFSS. In the absence of other potential confounders used in model 1, alcohol consumption and smoking were significantly associated with CHI (Table 6). The assumption that all states have the national level prevalence of CHI may yield inaccurate state-specific estimates because risk of CHI more than likely varies by state.

Therefore, model 3 is important because it may be duplicated to calculate state level estimates such as BRFSS datasets and to determine the state level prevalence of CHI.

The values of C-statistic for the three models in this study were 0.94, 0.92 and 0.88 respectively. Such high c-statistic values indicate stronger models with higher predictive accuracy of CHI in persons with associated demographic and risk factor variables in respective models. Even with the limited number of risk factors modeled after the BRFSS datasets, c-statistic is 0.88 indicating good prediction of the model. Thus, our data can serve as validation of the model based on BRFSS datasets.

Our study comes with several limitations as stated below,

1. We did not include participants of ages >/= 60 years due to the time and resource limitations for conducting this thesis study. This limits us from comparing the prevalence estimates and associations with CHI in this age group (minus the drug questions and life-time sex-partner questions which are not asked for this population by NHANES) to the results in previously published studies (4).

2. Since NHANES does not ask drug and lifetime sex-partner questions in adults of age >/= 60, we do not know the total true estimates of CHI prevalence and risk factor associations in adults inclusive >/= 60 years. In future, it will be useful to conduct similar studies in the baby boomers' cohort who are 49-69 years of age in 2014 because this cohort by itself is an important risk factor for CHI.

3. We did not present anti-HCV data in this study due to the time and resource limitations in performing the thesis work. National level anti-HCV prevalence estimates and its risk factor associations will enable to understand the larger picture of the CHI such as the conversion of the number of acute HCV cases into chronic HCV cases.

4. As mentioned in earlier studies (4, 29, 41), results from NHANES data are only applicable to the non-institutionalized U.S. civilian population, which underestimates the true prevalence of CHI because of the exclusion of incarcerated, homeless and institutionalized population in the datasets. Results from this thesis cannot be extrapolated to the aforementioned high-risk groups for whom the prevalence of CHI is likely to be higher.

5. Questionnaire data is relied on self-reporting and therefore subject to recall bias. Use of IDU and other drugs, having >/= 10 lifetime sex-partners are socially stigmatized activities which may result in participants being unwilling to admit to this behavior resulting in an underestimation of these factors in CHI prevalence and risk factor associations.

CONCLUSIONS

We conclude that the estimated persons infected with CHI as of 2014 is approximately 2.2 million in the civilian non-institutionalized U.S population sampled by NHANES. It has somewhat declined since 2010 which may be because of HCV related mortality, however the true prevalence estimates of CHI will be significantly higher when incarcerated, homeless and other population excluded from NHANES are included in the analysis. Injection drug use continues to be the strongest risk factor for CHI. Persons with two or more substance abuse risk factors have the highest odds of getting CHI. Results from this study will be critical in development of public health policies and guidelines for the identification of underappreciated CHI population and linking them to appropriate treatment and care.

REFERENCES:

1.   Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989;244(4902):359-62.

2.   Alter MJ, Margolis HS, Krawczynski K, Judson FN, Mares A, Alexander WJ, et al. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. N Engl J Med. 1992;327(27):1899-905.

3.   Schaefer EA, Chung RT. Anti-hepatitis C virus drugs in development. Gastroenterology. 2012;142(6):1340-50 e1.

4.   Denniston MM, Jiles RB, Drobeniuc J, Klevens RM, Ward JW, McQuillan GM, et al. Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003 to 2010. Ann Intern Med. 2014;160(5):293-300.

5.   Lingala S, Ghany MG. Natural History of Hepatitis C. Gastroenterol Clin North Am. 2015;44(4):717-34.

6.   Volk ML, Tocco R, Saini S, Lok AS. Public health impact of antiviral therapy for hepatitis C in the United States. Hepatology. 2009;50(6):1750-5.

7.   Kuo G, Choo QL, Alter HJ, Gitnick GL, Redeker AG, Purcell RH, et al. An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science. 1989;244(4902):362-4.

8.   Simmonds P. Variability of the hepatitis C virus genome. Curr Stud Hematol Blood Transfus. 1998(62):38-63.

9.   Irshad M, Ansari MA, Singh A, Nag P, Raghvendra L, Singh S, et al. HCV-genotypes: a review on their origin, global status, assay system, pathogenecity and response to treatment. Hepatogastroenterology. 2010;57(104):1529-38.

10.   National Institutes of Health Consensus Development Conference statement Management of Hepatitis C: 2002 June 10-12, 2002. HIV Clin Trials. 2003;4(1):55-75.

11.   Vieyres G, Thomas X, Descamps V, Duverlie G, Patel AH, Dubuisson J. Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. J Virol. 2010;84(19):10159-68.

12.   Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med. 2001;345(1):41-52.

13.   Lu M, Li J, Rupp LB, Zhou Y, Holmberg SD, Moorman AC, et al. Changing trends in complications of chronic hepatitis C. Liver Int. 2017.

14.   El-Serag HB, Kramer J, Duan Z, Kanwal F. Epidemiology and outcomes of hepatitis C infection in elderly US Veterans. J Viral Hepat. 2016;23(9):687-96.

15.   Kanwal F, Kramer JR, Ilyas J, Duan Z, El-Serag HB. HCV genotype 3 is associated with an increased risk of cirrhosis and hepatocellular cancer in a national sample of U.S. Veterans with HCV. Hepatology. 2014;60(1):98-105.

16.   Maasoumy B, Wedemeyer H. Natural history of acute and chronic hepatitis C. Best Pract Res Clin Gastroenterol. 2012;26(4):401-12.

17.   Younossi Z, Blissett D, Blissett R, Henry L, Younossi Y, Beckerman R, et al. In an era of highly effective treatment, hepatitis C screening of the United States general population should be considered. Liver Int. 2017.

18.   Davis KL, Mitra D, Medjedovic J, Beam C, Rustgi V. Direct economic burden of chronic hepatitis C virus in a United States managed care population. J Clin Gastroenterol. 2011;45(2):e17-24.

19.   Razavi H, Elkhoury AC, Elbasha E, Estes C, Pasini K, Poynard T, et al. Chronic hepatitis C virus (HCV) disease burden and cost in the United States. Hepatology. 2013;57(6):2164-70.

20.   Goolsby Hunter A, Rosenblatt L, Patel C, Blauer-Peterson C, Anduze-Faris B. Clinical characteristics, healthcare costs, and resource utilization in hepatitis C vary by genotype. Curr Med Res Opin. 2017;33(5):829-36.

21.   Nkontchou G, Ziol M, Aout M, Lhabadie M, Baazia Y, Mahmoudi A, et al. HCV genotype 3 is associated with a higher hepatocellular carcinoma incidence in patients with ongoing viral C cirrhosis. J Viral Hepat. 2011;18(10):e516-22.

22.   Probst A, Dang T, Bochud M, Egger M, Negro F, Bochud PY. Role of hepatitis C virus genotype 3 in liver fibrosis progression--a systematic review and meta-analysis. J Viral Hepat. 2011;18(11):745-59.

23.   El Khoury AC, Vietri J, Prajapati G. The burden of untreated hepatitis C virus infection: a US patients' perspective. Dig Dis Sci. 2012;57(11):2995-3003.

24.   Rosenthal ES, Graham CS. Price and affordability of direct-acting antiviral regimens for hepatitis C virus in the United States. Infect Agent Cancer. 2016;11:24.

25.   Guerra J, Garenne M, Mohamed MK, Fontanet A. HCV burden of infection in Egypt: results from a nationwide survey. J Viral Hepat. 2012;19(8):560-7.

26.   Hepatitis C--global prevalence (update). Wkly Epidemiol Rec. 1999;74(49):425-7.

27.   Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. N Engl J Med. 1999;341(8):556-62.

28.   Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. Ann Intern Med. 2006;144(10):705-14.

29.   Edlin BR, Eckhardt BJ, Shu MA, Holmberg SD, Swan T. Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology. 2015;62(5):1353-63.

30.   Grebely J, Prins M, Hellard M, Cox AL, Osburn WO, Lauer G, et al. Hepatitis C virus clearance, reinfection, and persistence, with insights from studies of injecting drug users: towards a vaccine. Lancet Infect Dis. 2012;12(5):408-14.

31.   Garfein RS, Doherty MC, Monterroso ER, Thomas DL, Nelson KE, Vlahov D. Prevalence and incidence of hepatitis C virus infection among young adult injection drug users. J Acquir Immune Defic Syndr Hum Retrovirol. 1998;18 Suppl 1:S11-9.

32.   Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. Centers for Disease Control and Prevention. MMWR Recomm Rep. 1998;47(RR-19):1-39.

33.   Garfein RS, Vlahov D, Galai N, Doherty MC, Nelson KE. Viral infections in short-term injection drug users: the prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses. Am J Public Health. 1996;86(5):655-61.

34.   Smith BD, Morgan RL, Beckett GA, Falck-Ytter Y, Holtzman D, Teo CG, et al. Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945-1965. MMWR Recomm Rep. 2012;61(RR-4):1-32.

35.   Smith BD, Yartel AK, Krauskopf K, Massoud OI, Brown KA, Fallon MB, et al. Hepatitis C virus antibody positivity and predictors among previously undiagnosed adult primary care outpatients: cross-sectional analysis of a multisite retrospective cohort study. Clin Infect Dis. 2015;60(8):1145-52.

36.   Finelli L, Miller JT, Tokars JI, Alter MJ, Arduino MJ. National surveillance of dialysis-associated diseases in the United States, 2002. Semin Dial. 2005;18(1):52-61.

37.   Fissell RB, Bragg-Gresham JL, Woods JD, Jadoul M, Gillespie B, Hedderwick SA, et al. Patterns of hepatitis C prevalence and seroconversion in hemodialysis units from three continents: the DOPPS. Kidney Int. 2004;65(6):2335-42.

38.   Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National health and nutrition examination survey: sample design, 2011-2014. Vital Health Stat 2. 2014(162):1-33.

39.   Taylor AL, Denniston MM, Klevens RM, McKnight-Eily LR, Jiles RB. Association of Hepatitis C Virus With Alcohol Use Among U.S. Adults: NHANES 2003-2010. Am J Prev Med. 2016;51(2):206-15.

40.   Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol. 2012;12:82.

41.   Ditah I, Ditah F, Devaki P, Ewelukwa O, Ditah C, Njei B, et al. The changing epidemiology of hepatitis C virus infection in the United States: National Health and Nutrition Examination Survey 2001 through 2010. J Hepatol. 2014;60(4):691-8.

42.   Ly KN, Xing J, Klevens RM, Jiles RB, Ward JW, Holmberg SD. The increasing burden of mortality from viral hepatitis in the United States between 1999 and 2007. Ann Intern Med. 2012;156(4):271-8.

43.   Smith BD, Morgan RL, Beckett GA, Falck-Ytter Y, Holtzman D, Ward JW. Hepatitis C virus testing of persons born during 1945-1965: recommendations from the Centers for Disease Control and Prevention. Ann Intern Med. 2012;157(11):817-22.

44.   VA M. Screening for Hepatitis C Virus Infection in Adults: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2013;159(5).