



W 4.5 A214s 2008  
Adams, John 1979-  
Sexual identity and  
allostatic load

UNTHSC - FW



M03INK

LEWIS LIBRARY  
UNT Health Science Center  
3500 Camp Bowie Blvd.  
Ft. Worth, Texas 76107-2699



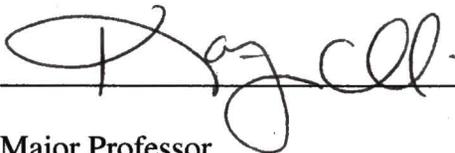
Adams, John P., Allostatic Load and Sexual Identity. Master of Public Health (Epidemiology), May 2008, 86pp., 4 tables, 1 figure, bibliography, 32 titles.

It is proposed that health disparities manifest in non-heterosexuals via stress experienced due to discrimination. To test this, allostatic load (conceptualized as stress-induced cumulative biological dysregulation) was compared between heterosexuals and non-heterosexuals by utilizing biomarker data from the 2001-2004 iterations of the National Health and Nutrition Examination Survey. Propensity score matching was implemented to increase group comparability and the association was reevaluated. Before and after propensity score matching there was no statistically significant association between sexual identity and having an allostatic load score (odds ratio = 1.202 95% confidence interval (0.663, 2.181) and 1.299 (0.638, 2.646), respectively); however, differences were found in HDL cholesterol and glycohemoglobin profiles. Allostatic load may be an inappropriate measure for non-heterosexuals as more sensitive biomarkers may be needed.

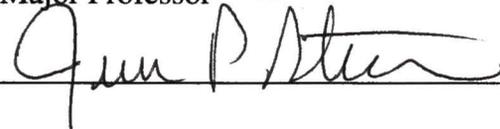
SEXUAL IDENTITY AND ALLOSTATIC LOAD

John Adams, BS

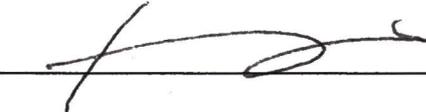
APPROVED:

  
\_\_\_\_\_

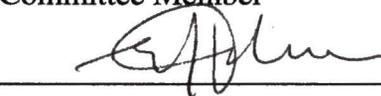
Major Professor

  
\_\_\_\_\_

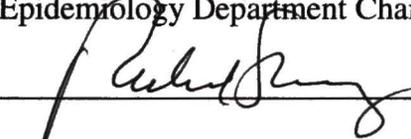
Committee Member

  
\_\_\_\_\_

Committee Member

  
\_\_\_\_\_

Epidemiology Department Chair

  
\_\_\_\_\_

Dean, School of Public Health

**SEXUAL IDENTITY AND ALLOSTATIC LOAD**

**THESIS**

**Presented to the School of Public Health  
University of North Texas  
Health Science Center at Fort Worth  
in Partial Fulfillment of the Requirements**

**for the Degree of**

**Master of Public Health**

**By**

**John Adams, BS**

**Fort Worth, Texas**

**May, 2008**

## ACKNOWLEDGEMENTS

I would like to thank Dr. Kathryn Cardarelli for tirelessly supporting my research. Without her guidance, none of this would be possible. Many thanks are also owed to my committee members, Dr. Jim Stimpson, for guiding the conceptualization and providing a framework for this project and Dr. Fang Fang Zhang, for providing instruction in epidemiologic data analysis and always giving me support and advice.

Additionally, I would like to acknowledge Kelly Ylitalo, Elisa Priest, David Nicewander, Dr. Ed DeVol and Dr. Giovanni Filardo for providing feedback throughout this research project.

## TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
Chapter	
1. INTRODUCTION	1
1.1 Summary	1
1.2 Problem / Hypothesis	2
1.3 Significance	4
2. LITERATURE REVIEW	7
2.1 Allostatic Load in Marginalized Minority Populations	7
2.2 Prevalent Conditions and Health Burden in the Gay/Lesbian Community	10
2.3 Effects of Concealing Sexual Identity	14
2.4 Effects of Minority Stress and Discrimination	15
3. METHODS	16
3.1 Treatment of Demographics	18
3.2 Health Indices Treatment	20
3.3 Allostatic Load	23
3.4 Propensity Score Matching	27
4. RESULTS	29
4.1 Key demographic	29
4.2 Health Indicator Comparisons	31
4.3 Allostatic Load	33
4.4 Health Indicator Comparisons and Allostatic with the Application of Propensity Score Matching	34
5. DISCUSSION	36
6. LIMITATIONS	45
REFERENCES / BIBLIOGRAPHY	49
APPENDIX	54

## LIST OF TABLES

	Page
TABLE 1. Key Demographic Characteristics across Sexual Identity, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module.	55
TABLE 2. Percent of Participants with Clinically Significant Values of Individual Indices by Sexual Identity, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module.	56
TABLE 3. Key Demographic Characteristics Across Sexual Identity after Propensity Score Matching, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module.	57
TABLE 4. Odds Ratio (and 95% CI) of Having a Clinically Significant Index Value Among Non-Heterosexuals as compared to Heterosexuals With and Without Matching, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module in Specific Biologic Indices.	58

## LIST OF FIGURES

	Page
FIGURE 1. Conceptual model relating sexual identity to allostatic load	59

## CHAPTER 1

### INTRODUCTION

#### *1.1 Summary*

In this study, allostatic load is employed as an indicator of physical health among participants in the continuous National Health and Nutrition Examination Survey (NHANES IV), 2001-2004 examination cycles. In addition to utilizing a scoring mechanism to measure cumulative dysregulation (allostatic load) across biologic systems, propensity score matching is performed to simulate the effects of randomization in order to minimize residual confounding.

Allostatic load is assessed by creating an index using clinical cut-points for biological markers representing several domains of physical health. For clinically significant values of each biomarker, a score of 1 is assigned, with a default score of 0. These values are summed to create an allostatic load score. Logistic regression is used to compare those individuals with allostatic load scores to those with no allostatic load scores between participants of different sexual identities while correcting for relevant factors found in current publications assessing allostatic load, physical health among gay men, lesbians, and bisexuals and/or sexual identity. Similarly propensity score matching is performed (based on factors identified in the literature that are associated with sexual identity) in order to increase comparability between heterosexuals and non-heterosexuals and control for confounding.

## *1.2 Problem/Hypothesis*

Gay men and lesbians are sometimes discriminated against internally (by other gay men and lesbians) and externally (by heterosexuals) (Skidmore, Linsenmeier, and Bailey, 2006) and it is shown that minority groups are subject to negative health outcomes when compared to non-minority groups (Chu, Miller, and Springfield, 2007). Allostatic load has arisen as one way of explaining the increase in mortality/morbidity in select minority populations. Allostatic load, or stress-induced biologic dysregulation, is cumulative in its effect and most greatly impacts adult health, exclusive of the elderly (Seeman, Singer, Ruff, Love, & Levy-Storms, 2002; & Allsworth, Weitzen, and Boardman, 2005). Gay men, lesbians and bisexuals have a higher prevalence of certain chronic diseases, and it is possible that allostatic load may be one route through which stress (due to discrimination) influences chronic health outcomes in an adult homosexual / bisexual population. It is then likely the case that those who identify as homosexual or bisexual may have higher allostatic load scores than those who identify as heterosexual due to the inherent stress of being gay / bisexual. Many exposures that may affect allostatic load, such as unhealthy lifestyle behaviors, are associated with sexual identity (Ungvarski & Grossman, 1999). This may confound the association between sexual identity and allostatic load score. In an attempt to increase comparability between heterosexuals and non-heterosexuals groups, propensity score matching is employed to accurately test the effect of sexual identity upon allostatic load. After propensity score matching, the effect of sexual identity upon allostatic load more accurately reflects the counterfactual construct, giving a less biased estimate of allostatic load by reducing

residual confounding. Sample size may not allow adequate power to detect small differences in measures of association. Thus, depending on sample size and magnitude of measures of association, the benefits of propensity score matching may prove inconclusive. The association between allostatic load and sexual identity is interpreted in light of these issues.

The specific aims of this study are to determine if there is an association between the stress experienced due to sexual identity and allostatic load and to determine if any associations exist between sexual identity and individual biomarker indices of health. The operational hypothesis is that gay men, lesbians, and bisexuals have higher allostatic load scores than heterosexuals and that differences in biomarker indices are evident when comparing values for heterosexual and non-heterosexuals.

### *1.3 Significance*

Research on gay and lesbian participants has yielded multiple findings which suggest that homosexuals are at increased risk of physical and mental health problems, including depression and HIV incidence. Although very little has found its way to publication in mainstream public health journals, it is believed that gay men, lesbians and bisexual have unique health concerns from heterosexual individuals. The Gay and Lesbian Medical Association cites several risk factors for negative health outcomes that are more prevalent in gay men and lesbians than heterosexuals. These risk factors include obesity, smoking, and chronic stress in lesbians and tobacco, drug and alcohol use, chronic stress and body image problems in gay men (Gay and Lesbian Medical Association, N.D.). This suggests that gay men and lesbians may have differing medical needs from heterosexual men and women. The stress experience of homosexuals and bisexuals as it relates to health outcomes and differing medical needs remains largely unstudied.

Additionally, because of the high prevalence of HIV in the gay community, research has been conducted on how stress affects HIV-related health outcomes, much to the exclusion of non-HIV related health outcomes. Considering the disproportionate burden of chronic health conditions in the homosexual / bisexual community and the current atmosphere in which gay men, lesbians, and bisexuals are afforded less social acceptance than heterosexuals, research focusing on sexual identity based stress as it is related to poor health outcomes in the non-heterosexual community becomes topical and necessary.

Recent publications have used allostatic load as a measure of biologic dysregulation by utilizing biomarkers representing several regulatory systems as a proxy measure. An allostatic load score (utilizing the same or similar biomarkers) can be created in a representative sample of the United States population to represent the cumulative physical effects of stress (allostatic load) associated with discrimination experienced by minority group, in this case gay men and lesbians.

Using NHANES biological data, this study attempts to fill a gap in the current gay/lesbian/bisexual health literature. Additionally (because heterosexuals and non-heterosexuals are different in terms of their health behaviors (Ungvarski & Grossman, 1999)) to increase group comparability, propensity score matching will be implemented. Propensity score matching is a developing methodology that attempts to match research participants based on their propensity of exposure (predicted by covariates of exposure). It seeks to simulate the effects of randomization and reduce or eliminate residual confounding (Oakes & Johnson, 2006). By reducing residual confounding, a more accurate representation of how sexual identity affects allostatic load may be seen. Heterosexuals and non-heterosexuals are shown to be quite different in terms of their health behaviors and other demographics, thus propensity score matching may prove beneficial in increasing the ability to detect differences in sexual identity groups which can be attributed solely to the stress incumbent upon homosexuals and bisexuals.

Whereas previous studies have researched the effects of sexual identity based discrimination on health conditions, this study is unique in that it seeks to determine the relationship between stress due to discrimination (based on sexual identity) and a

comprehensive, multi-system indicator of health (allostatic load) that may measure the deterioration of health before symptoms manifest clinically.

If non-heterosexuals are shown at present to be more likely to have a higher allostatic load values than heterosexuals, it suggests that chronic health conditions may be more prevalent in non-heterosexuals in the near future. Additionally, it may explain current health disparities between heterosexuals and non-heterosexuals. This body of research holds significance for gay, lesbian, and bisexual community groups, policy makers, public health entities and anyone who may act on behalf of the welfare of gay men, lesbians and bisexuals.

## CHAPTER 2

### LITERATURE REVIEW

#### *2.1 Allostatic load in marginalized minority populations*

Stress resulting from discrimination affects gay men and lesbians as well as many other marginalized groups. Sociologic minority stress (stress resultant from being disadvantaged in terms of social status, education, wealth or power) has been widely studied, and shown to affect both mental and physical health in minority populations. The mechanism by which sociologic minority stress impacts health, however, is less clear. Although it is shown that environment (both social and physical) is related to health outcomes, no one pathway has been elucidated that adequately explains this relationship. Bruce McEwen (2001) has introduced allostasis as a candidate pathway through which chronic stress may affect health. Allostatic load (a measure of allostasis) is the cumulative physiologic dysregulation of the body's systems due to adaptation to a changing environment (McEwen, 2001). This is influenced by the interaction of lifestyle, genetics, early and late life-course experiences and development. Allostatic load can be modified by diet, sleep, smoking, alcohol consumption and exercise. In general, the body responds to stress by secreting hormones that, in the short term, are beneficial to adapting to a stressor. However, in the long term, if secreted in excess these hormones promote ill effects to the body (McEwen, 2001). Long-term low socioeconomic position, a chronic stressor, may be a factor in mortality and early decline (Seeman, Crimmins, Huang, Singer, Bucur, Gruenewald, Berkman, and Reuben, 2004). It may follow that the

frequently-recurring stressor of discrimination on the basis of sexual identity may influence early decline in lesbians, bisexuals and gay men.

Historically, gay and bisexual men have had the stress of high HIV burden (in 1995 this group constituted 42% of all AIDS cases), an increased use of alcohol and drugs (Ungvarski & Grossman, 1999) which may be an escape/coping mechanisms for stress, and body image problems. In addition, homosexuals/bisexuals have to conceal their sexuality from both friends and family to differing degrees in multiple situations (Cole, Kemeny, Taylor, and Visscher, 1996; & Perez-Benitez, O'Brien, Carels, Gordon, and Chiros, 2007) to appear more acceptable and avoid discrimination, mental assault, and physical assault. Although concealment of sexual identity may decrease the probability of discrimination, concealment of sexual identity may also increase the intensity of the stress experienced by gay men, lesbians and bisexuals as a result of discrimination. Stress can then act to dysregulate the body as previously described by McEwen (2001). Stress, concealment of sexual identity, unhealthy lifestyle behaviors (alcohol/drug use), and an increased occurrence of health problems (Cochran & Mays, 2007) make the study of allostatic load within the gay, lesbian and bisexual community relevant.

It is proposed here that being non-heterosexual results in discrimination, which increases stress. The stress of discrimination may cause some to conceal their sexual identity, which may actually increase the stress experienced due to discrimination. As the stressor (discrimination) persists, stress hormones persist in the body, which affect biomarker indices of health. These biomarkers include metabolic markers

(glycohemoglobin, waist-to-hip ratio and cholesterol), cardiovascular markers (blood pressure and heart rate), and inflammatory markers (albumin and c-reactive protein). Allostatic load is directly affected by dysregulated biomarker indices and increased morbidity/mortality follows (as shown in figure 1). This model guides the literature reviewed which focuses on discrimination, stress, concealment and morbidity in the gay/lesbian/bisexual community.

## *2.2 Prevalent conditions and unhealthy behavior in the gay/lesbian community.*

Allostatic load may be influenced by lifestyle factors, such as smoking, excessive alcohol consumption and illegal drug use (McEwen, 2001). Additionally, chronic health conditions (such as those represented in excess in non-heterosexuals (Diamant & Wold, 2003; Cochran & Mays, 2007) may indicate higher allostatic load in those people who experience chronic health condition. Because there is evidence that non-heterosexuals have an excess of certain health conditions and an excess of unhealthy behaviors such as smoking, alcohol consumption, and illegal drug use (Ungvarski & Grossman, 1999), the study of allostatic load is especially appropriate for this community.

Although only a modest number of research projects have assessed gay and lesbian health, exclusive of HIV, it is recognized by the Gay and Lesbian Medical Association that gay men, bisexuals and lesbians have increased prevalence of unhealthy behaviors and health risk factors when compared to heterosexuals. These risk factors/behaviors include obesity, smoking, and chronic stress in lesbians, and smoking, drug use, alcohol use, chronic stress and body image problems in gay men (Gay and Lesbian Medical Association, N.D.). Additional evidence of excess unhealthy behavior as it is associated with sexual identity exists in asthma studies using pooled data from the National Health Interview Survey. Heck and Jacobson (2006) studied the prevalence of asthma among gay men and lesbians in same-sex relationships. They found that 13.5% of men and 14.3% of women in same-sex relationships as compared to 7.6% of men and 10.2% of women in opposite-sex relationships experienced asthma. It was also found that those people in same-sex relationships had a higher prevalence of smoking, stress

and obesity (in women). The adjusted odds ratios of having asthma in men and women in same-sex relationships as compared to those in opposite-sex relationships were 1.51 and 2.48, respectively, and may have been due to the increased prevalence of asthma risk factors in homosexuals.

Gay men, bisexuals and lesbians suffer a disproportionate burden of chronic physical health conditions when compared to heterosexuals. The 1999 Los Angeles County Health Survey of women aged 18 to 64 showed among 4023 heterosexual, 69 bisexual and 43 homosexual women, those that did not identify as heterosexual were associated with increased rates of both poor mental health days (7.4 in lesbians versus 4.2 days in heterosexuals,  $p < 0.01$ ) and poor physical health days (5.8 in bisexual women versus 3.6 days in heterosexuals,  $p < 0.05$ ) in the month preceding the study as well as increased report of diagnosis of heart disease (4.5% in heterosexuals, 18.6% in lesbians, and 11.6% in bisexual women,  $p < 0.001$ ). Lesbians were more likely than heterosexual women to use antidepressants among those with depressive disorder (OR 13.7, 95% CI: 1.5, 125.6) (Diamant and Wold, 2003). Additionally a cross-sectional study of 13 family practices in London including 934 heterosexual women, 373 heterosexual men, 38 gay men, 23 bisexual men, 26 lesbians and 85 bisexual women showed that gay men were more likely to report higher levels of psychological symptoms (OR 2.48, 95% CI 1.05-5.90), bisexual women were more likely to abuse alcohol (OR 2.73, 95% CI 1.70-4.40), bisexual men (OR 2.48, 95% CI 1.04-5.86) and women (2.53, 95% CI 1.60-4.00) and lesbians (3.13, 95% CI 1.41-6.97) were also shown to be more likely to smoke (King and Nazareth, 2006).

In the California Quality of Life Survey, gay men and lesbians were found to have unique health concerns as compared to heterosexual men and women. It was shown that gay men were more likely to report urinary tract problems (OR 3.93 95% CI 1.39 – 11.12), headache/migraine (OR 2.74 95% CI 1.43-5.23) and enteritis, colitis and stomach ulcer (OR 3.60 95% CI 1.41 – 9.20) than heterosexual men, and bisexual women were more likely to have back problems (OR 2.39 95% CI 1.10 – 5.20), be disabled (OR 2.46 95% CI 1.21 – 5.00), and suffer stomach problems (OR 2.77 95% CI 1.06 – 7.22) (Cochran & Mays, 2007).

Adults are not the only group affected by having a non-heterosexual identity. A community-based cross-sectional study in two upper middle class high schools studied the health problems of gay, lesbian and bisexual high school students. This study shows that gay, lesbian, unsure (of sexual identity), and bisexual adolescents had increased risk in mental health ( $\chi^2 = 51.64$ ,  $p = 0.0001$ ) and general health ( $\chi^2 = 18.63$ ,  $p = 0.0001$ ) (Lock & Steiner, 1999) as measured by an anonymous questionnaire, the Juvenile Wellness and Health Survey, assessing health behavior. Risk was measured using mean scores within 5 categories addressing general health, mental health, and general risk taking among others and evaluating tests for association ( $\chi^2$ ). As stress acts over the lifespan (McEwen, 2001), it is important to note, as evidenced by Lock and Steiner (1999) that sexual identity affects health as early as high school.

These studies show an increased prevalence of poor health outcomes and heavier alcohol / tobacco / drug use in non-heterosexuals. Poor health outcomes may indicate increased allostatic load and substance abuse may increase allostatic load. This further

substantiates the need to assess allostatic load across sexual identities. This is also relevant as the health effects of discrimination resulting from being gay / lesbian / bisexual may begin to manifest as early as high school and continue to accumulate across the lifespan. However, it is also important to assess the stress experience of non-heterosexuals due to discrimination to make an effective argument for evaluating allostatic load across sexual identities.

### *2.3 Effects of Concealing Sexual Identity*

Although discrimination may directly affect stress, it may also cause some people to conceal their sexual identity to avoid discrimination. While this may make those people concealing more socially acceptable, there is mounting research demonstrating that the stress associated with concealment of one's sexual identity can affect physical health measures in homosexuals and bisexuals. Similarly, concealment induced by discrimination may also contribute to the allostatic load experience of non-heterosexuals.

Perez-Benitez, O'Brien, Carels, Gordon and Chiros (2007) studied the effect of disclosing homosexual orientation on cardiovascular outcomes. They showed in a study of 27 gay men that those with the highest concealment of their homosexuality who disclosed more information about their homosexuality in a controlled lab session had the highest levels of psychophysiologic recovery as compared to those that had low levels of disclosure of homosexuality (for stroke volume in 2 way ANOVA examination of interaction of disclosure and level of concealment,  $p = 0.037$ ). Investigators used this evidence to suggest that disclosure decreases the risk of cardiovascular disease and hypertension. The results are limited in that they focus mainly on cardiopulmonary measures alone to determine the health effects of non-disclosure as opposed to multiple indicators of health status across the entire physiologic spectrum. Perhaps the most compelling research showing elevated physical health risk in gay men who conceal their sexual identity comes from those HIV-negative men that participated in the Natural History of AIDS Psychology Study. It showed that incidence of infectious disease (non-STD / HIV) in those that concealed their sexuality over a 5-year follow up to be 2.91

times the incidence of those that did not conceal their sexual identity. An odds ratio of 3.18 when comparing the incidence of cancer of those concealing to those not concealing sexual identity was also documented in this study (Cole, Kemeny, Taylor, & Visscher, 1996). These associations remained statistically significant when controlling for drug use, anxiety, age, ethnicity, socioeconomic status, coping style, behavioral patterns and reporting of social desirability (odds ratio = 9.79).

In contrast, a study that investigated the association between work disclosure (of sexual orientation) and salivary cortisol levels was conducted on a cohort of 73 gay and bisexual men to determine if concealment of sexual identity relates to negative health effects (for which cortisol was a proxy measure). The study found cortisol levels were higher in those that had higher disclosure of sexual identity (Huebner & Davis, 2005). This is in contrast to the previous studies. High cortisol levels are indicative of emotional and physiologic activation (stress) (Huebner & Davis, 2005), and thus this study suggests that “out” gay people may have increased risk for future health problems as compared to their “closeted” counterparts. This presents a challenge to the classic exposure-outcome relationship implying that many factors (including context of concealment) may act as modifiers in the association between level of disclosure and health. For example, the arena of their life in which homosexuals/bisexuals choose to conceal their orientation may influence the degree to which concealment affects their health.

## *2.4 Effects of Minority Stress and Discrimination*

Disclosure, however, only constitutes one mechanism through which discrimination may cause sexual identity-related stress and influence negative health outcomes. Other studies examine the effect of discrimination in general. Discrimination directly resulting in chronic stress is the most direct route through which sexual identity-based discrimination may result in an increase in allostatic load.

The 1989 CARDIA study of black and white men and women revealed that 33, 39, 52 and 56 percent of homosexually-experienced black women, black men, white women and white men, respectively, experienced sexual orientation-based discrimination. In addition, over 75 percent of black men and women experienced racial discrimination and almost 90 percent of women experienced gender discrimination (Krieger & Sidney, 1997). The stress of dealing with the discrimination described above is believed by some to be part of the gay/lesbian experience. The Gay and Lesbian Medical Association (N.D.) presents chronic stress as an issue that gay men and lesbians should discuss with their medical providers. Additionally, Huebner and Davis (2006) sought to describe the relation between discrimination and physical health in a group of 361 gay and bisexual men. Outcomes included number of physician visits, sick days from work, and both prescription and non-prescription medication use. The study measured perceived discrimination as an exposure using internet, telephone and self-administered surveys. The results indicated a clear relationship between perceived discrimination and physical health among those that had attained a higher level of education (as measured in years of formal education completed). Among those men with

lower levels of education a U-shaped relationship was noted. This is similar to the results of studies of perceived discrimination and health outcomes in African American men, where education is substituted for occupational status as a measure of socioeconomic position (Huebner and Davis, 2006). This evidence suggests not only that discrimination is present and affects the health of homosexuals and bisexuals, but that established models for other minority groups should be tested in gay and lesbian populations. This further supports the idea of testing allostatic load across sexual identity categories.

## CHAPTER 3

### METHODS

These data are abstracted from the National Health and Nutritional Examination Survey (NHANES IV) continuous modules, 2001-2002 and 2003-2004, which was conducted by the National Center for Health Statistics (NCHS). This comprises four cycles of administration, with approximately 5,000 people surveyed per administration per year. NCHS produces vital and health statistics for the United States and is part of the Centers for Disease Control and Prevention. This survey is conducted in 15 various counties every year. It includes both an interview and physical examination (containing demographic, socioeconomic, dietary, laboratory, dental and other biometric measurements conducted by medical professionals). Mobile examination centers are utilized for laboratory and physical examination. Tests are dependent upon age of the participant, with more data being collected for the older participants. NHANES utilizes a “complex, multistage, probability sampling design (Centers for Disease Control and Prevention, 2007).” NHANES over-samples those participants that are 60 years of age and older, Hispanic and African American. The sampling scheme is comprised of four stages: selecting primary sampling units (counties or groups of counties) (PSUs), segmenting those PSUs with probability proportional to a measure of size, listing and drawing households from the segmented PSUs, and listing and drawing individuals from these households. This is designed for the purposes of having a representative sample, after weighting is applied. This study is approved by the University of North Texas Health Science Center’s Institutional Review Board, and the data are publicly available

and no identifiers are included. This research poses no risk of any breach of confidentiality.

Demographic data come from interviews conducted by trained personnel. Biomarkers are all taken from blood draws and physical examination. When weighting is applied the data are nationally representative; however, due to sample size constraints resultant to the small number of gay/lesbian/bisexual participants it may be the case given that weighting will not be appropriate after propensity score matching. Standard error will be evaluated on weighted analysis to address this. Also due to sample size constraints, it will be necessary to analyze and discuss power after analysis.

### 3.1 *Treatment of demographics*

Sexual identity includes heterosexual, homosexual, bisexual, and unsure. It is measured by the item, “How you describe your orientation” in the sexual behavior section of the NHANES 2001-2002 and 2003-2004 questionnaire modules. In modeling the data, homosexual and bisexual participants are combined into one group and compared against heterosexuals as the referent group. Sexual identity is grouped separately for men and women in this dataset, thus they are combined creating one category for both men and women for sexual identity.

Demographic variables of interest include level of education, age, ethnicity/race, income and gender and their categorization is based on Seeman, et al.’s 2008 study of allostatic load utilizing NHANES III data. Values are reported separately for heterosexuals and non-heterosexuals. Age is reported both as mean age (in years) as well as percent of participants within 10 year age categories. Level of education is reported as the percent that have completed: less than high school (0-12 years), 12 years of education or equivalent (high school graduates or GED), and more than high school (12+ years). Race/ethnicity is reported together as one variable with “Mexican American,” “white,” “black,” and “other” as categories. These data are available already combined as race/ethnicity in the NHANES continuous modules. It is reported as the percentage of each race/ethnicity within each sexual identity category. Household income is reported as poverty income ratio (PIR) and categorized as <1.0, 1.0-1.99, 2.0-2.99, 3.0-3.99, 4.0-4.99 and 5.0+. Poverty income ratio is the ratio of the participant’s income to the

national poverty threshold as set by the United States Census Bureau. Gender is reported as percent female per sexual identity.

Covariates for multivariate analysis include “more than moderate” alcohol use, tobacco use, street drug use, and self-rated health. Alcohol use is reported as the weighted percentage of those who had more than 2 alcoholic drinks per day in men and more than 1 alcoholic drink a day for women on those days they drank alcohol for each sexual identity category. This definition of moderate drinking comes from the United States Department of Agriculture and United States Department of Health and Human Services (2005) definition of moderate drinking. These entities define moderate drinking as up to 1 drink per day for women and up to 2 drinks per day for men. Any values above this will be classified as “more than moderate.” This is classified as such due to the inability of dichotomous drinking categories (e.g. drinkers and non-drinkers) to be informative in terms of health behavior. Comparing those drinking more than moderate amounts to those who drink moderately may be more reflective of health behavior, as drinking moderate amounts or less than moderate amounts is not advised against in a healthy population. Smoking status is categorized into current, former and never smokers by the questions, “Have you smoked at least 100 cigarettes in your life,” and “Do you now smoke cigarettes?” Those responding positively to both questions are classified as current smokers, whereas those that respond negatively to both are classified as non-smokers. Those indicating they had smoked at least 100 cigarettes, but are not currently smoking are classified as former smokers. This treatment of smoking behavior is consistent with Seeman’s (2008) study utilizing NHANES III data. Illegal drug use is

reported as the weighed percent that have ever used street drugs, excluding marijuana, in each sexual identity category. In health literature, street drug use is not consistently quantified, but there are clear differences on ever having used illegal drugs when comparing heterosexuals to non-heterosexual. The allostatic load literature reviewed for this study does not address illegal drug use. The lack of availability of more quantitative measures (dose, duration, etc.) of illegal drug use in both 2001-2002 and 2003-2004 NHANES modules along with lack of clear guidance from allostatic load literature led to the selection of this categorization method. Current health status is measured as weighted percent of those who responded, “excellent,” “very good,” “good,” “fair,” or “poor” to the question, “Would you say your health in general is...” This measure was selected because some studies measuring health difference in heterosexuals and non-heterosexuals use perceived health status measures such as poor mental health day (Diamant and Wold, 2003).

Chi-square analysis is conducted to test homogeneity of demographics and covariate across sexual identity for categorical data. For mean age, an independent t-test is conducted to determine if there is a difference in mean values for age between sexual identity categories. Those covariate categories that show a lack of homogeneity or difference in means are used to propensity score match upon along with demographic data. This is further discussed later in section 2.4.

### 3.2 Health indices treatment

To determine if sexual identity affects allostatic load scores, it is first necessary to select biologic indicators to create an allostatic load variable. This is accomplished by abstracting: diastolic blood pressure (mm Hg), systolic blood pressure (mm Hg), glyco-hemoglobin (%), c-reactive protein (CRP) (mg/dL), albumin (g/dL), HDL cholesterol (mg/dL), total cholesterol (mg/dL), resting heart rate (beats per minute) and BMI and assigning a value of 1 for high risk values of each biomarker (other values will be scored 0). The following values are used for cut-points indicating clinical significance: albumin <3.8 g/dL, CRP  $\geq$  0.3 mg/dL, Total cholesterol  $\geq$  240 mg/dL, HDL cholesterol < 40 mg/dL, glyco-hemoglobin  $\geq$  6.4 %, resting heart rate  $\geq$  90 beats per minute, systolic blood pressure  $\geq$  140 mm Hg, and diastolic blood pressure  $\geq$  90 mm Hg (Seeman, Merkin, Crimmins, Koretz, Charette & Karlamangla, 2008). These are the standard measures used in other studies and are used here for consistency. BMI was chosen as  $\geq$  30 based on CDC guidelines of obesity (2007). Other studies use waist-hip ratio or BMI, with waist-hip ratio being the preferential measure. Waist-hip data are unavailable for NHANES continuous modules, thus BMI was chosen. BMI is calculated from weight and height data and is readily available in NHANES continuous modules. The numerical formula is  $BMI = (\text{weight (pounds)} * 703) / \text{height (inches)}^2$ ; however, BMI is provided in NHANES continuous modules already calculated. Heart rate is calculated from 30 second heart rate by multiplying by two. Blood pressure readings are the average of 3 blood pressure readings for each individual. Where any one reading is missing the participant is excluded. All other biomarkers are taken from blood based laboratory

analysis. If any one biomarker is missing that observation was not included in the analysis of allostatic load score. Allostatic load should be measured across multiple systems of the body, and these biomarkers represent cardiovascular health (blood pressure, HDL cholesterol, and heart rate), metabolic health (BMI, glyco-hemoglobin, and cholesterol) and inflammation (CRP and albumin).

Logistic regression is used to test whether each dichotomous biomarker variable is associated individually with sexual identity. The outcome is given a score of 1 in a dichotomized biomarker group using the non-heterosexual group as the exposed category. The measure of association for this analysis is an odds ratio, with 95% confidence intervals. Specific models are created for each index based on relevant factors that predict clinically significant values.

Smoking, total cholesterol level, socio-economic position (SEP) (proxied by PIR) and BMI influence an individual's probability of having low albumin levels (Danesh, Collins, Appleby & Peto, 1998). BMI, SEP (proxied by PIR), ethnicity, smoking status, alcohol use, and gender may affect HDL cholesterol and total cholesterol levels (Centers for Disease Control and Prevention, 2007). Diabetes (for which glycohemoglobin level can be used as a proxy) is affected by BMI, age, race and total cholesterol (National Diabetes Information Clearinghouse, 2007). Modeling BMI against sexual identity will need to control for race and age (Centers for Disease Control and Prevention, 2007). CRP levels are influenced by BMI, smoking status, diabetes (proxied by glycohemoglobin values), smoking, and ethnicity (Danesh, Collins, Appleby & Peto, 1998). Blood pressure models should control for BMI, alcohol use, diabetes (proxied by

glycohemoglobin values), smoking status and race (Centers for Disease Control and Prevention, 2007). Heart rate is influenced by age, race, and high blood pressure (Mayo Clinic Staff, 2007). All of these factors are controlled for in the individual models for each biomarker index as it is influenced by sexual identity.

### *3.3 Allostatic Load*

For each participant all dichotomized biomarker values (i.e., 1 or 0 based on methods described in section 2.2) are summed to calculate an allostatic load score.

Previous studies using NHANES III have established a framework for the treatment of variables for constructing an allostatic load score (Allsworth, J. E., Weitzen, S., and Boardman, L. A., 2005; & Seeman, Merkin, et al., 2008). This is the established method in Seeman's (2004) work in the MacArthur studies of successful aging.

Logistic regression is used to model sexual identity against allostatic load.

Allostatic load is modeled such that those values 1+ are the outcome of interest, with those scoring 0 being the comparison group. Several variables can act as confounding variables in the relationship between sexual identity and allostatic load score. These include current health status, drug use, smoking status, and alcohol drinking behavior as they may be related to the outcome (high allostatic load) and the exposure (sexual identity). They are tested individually in the model and any one that changes the beta coefficient of sexual identity (modeled against allostatic load) by 10% or more is added to the model. Age is included in the model, along with poverty income ratio, education and race / ethnicity based on evidence from other studies (Seeman, Merkin, et al., 2008; & Geronimus, Hicken, Keene & Bound, 2006).

### *3.4 Propensity score matching*

As described above, gay men, lesbians and bisexuals may differ from heterosexuals in regard to their health behaviors. This can affect the analysis of these data to the extent to which these two groups are different. For example, if an association is seen between sexual identity and allostatic load it could simply be due to confounding from some external factor(s). Because these groups are so dissimilar in other studies in terms of health behavior, propensity score matching is implemented here to determine the effect of sexual identity on allostatic load in groups with comparable baseline characteristics.

Propensity score matching has arisen as a tool in social epidemiology to account for the lack of comparability in comparison groups in observational studies. The counterfactual framework, which is an exercise that examines the theoretical view of a causal effect by thinking of a comparison of the same group experiencing two different exposure experiences, is one way that epidemiologists examine group comparability. If, when examining two comparison groups, one can say that the comparator group is similar to the counterfactual, then comparability in groups is achieved. One way of achieving comparability, via the counterfactual framework, is propensity score matching. Propensity score matching achieves this by making comparison groups similar based on covariates of exposure. One method of actualizing propensity score matching is by logistic regression modeling. The covariates of exposure are modeled against an outcome of the dichotomous exposure (Oakes & Johnson, 2006).

The unexposed individuals are then matched to the exposed based on the probability of exposure from the logistic regression model. Usually this is achieved by matching the closest values of probability of exposure between unexposed and exposed groups within a range of +/- 0.01. After matching, both groups (the exposed and matched unexposed) should be comparable based on probability of exposure.

Propensity score matching is implemented here using logistic regression with sexual identity as an outcome. The demographic variables and covariates are used to predict the exposure. The probability of exposure based on the regression equation is output for each individual. The probability for each non-heterosexual individual is matched to its nearest probability estimate for one person in the heterosexual category within 0.01 calipers (unless the data require smaller caliper ranges for matching to be possible).

Analysis of non-heterosexuals then is conducted against a comparison group of propensity score matched heterosexuals. Demographic analysis, logistic regression analysis of dichotomous biomarker variables and of dichotomous allostatic load score are repeated for the propensity-score matched group.

All statistical analysis are performed using the SAS system, version 9.1.

## CHAPTER 4

### RESULTS

#### *4.1 Key demographics*

The sample size analyzed for demographics contained 5541 participants that responded to the sexual behavior module across the 2001-2004 survey years. Heterosexuals comprised 5374 of the respondents while 167 respondents identified as non-heterosexual (i.e., bisexual or homosexual). Participants ranged in age from 20 to 59 years of age. Table 1 summarizes weighted demographic figures. Age distribution did not show that data were heterogeneous across 10 year age categories, nor were there any statistically significant differences by education by race, PIR, gender or current health status at statistical significance level of 0.05. 27 percent of non-heterosexuals were 20-29 years of age, 31 percent were 30-39 years of age, 27 percent were 40-49 years of age, and 15 percent were 50-60 years of age. 25 percent of heterosexuals were 20-29 years of age, 25 percent were 30-39 years of age, 29 percent were 40-49 years of age, and 22 percent were 50-60 year of age. 76 percent of participants were white, 8 percent Mexican-American, 6 percent black and 10 percent identified as "other" among non-heterosexuals. While among heterosexuals 71 percent were white, 8 percent black, 12 percent Mexican-American and 10 percent were other races/ethnicities. In regard to income, 12 percent of non-heterosexuals had a PIR of < 1, 19 percent had a PIR of 1.0-1.99, 21 percent had a PIR of 2.0-2.99, 12.6 had a PIR of 3.0-3.99, 8 had a PIR of 4.0-4.99 and 27 percent had a PIR above 5. Heterosexuals were broken down as, 13 percent with a PIR < 1, 18 percent

having a PIR of 1.0-1.99, 15 having a PIR of 2.0-2.99, 15 having a PIR of 3.0-3.99, 13 having a PIR of 4.0-4.99 and 26 percent having a PIR above 5.

Education ( $p = 0.03$ ), drinking above moderation ( $p = 0.05$ ), ever using street drugs ( $p < 0.001$ ), and smoking status ( $p < 0.01$ ) were not homogenous across sexual identity categories. Whereas a similar amount of non-heterosexuals and heterosexuals completed less than high school (12 and 15 percent, respectively), 70 percent of non-heterosexuals completed more than high school and 60 percent of heterosexuals completed more than high school. Drinking behavior indicated that 65 percent of non-heterosexuals usually consumed levels above moderation when they drank as compared to 55 percent of heterosexuals. 45 percent of non-heterosexuals had used illegal street drugs other than marijuana, where only 21 percent of heterosexuals had done the same. Smoking behavior was also different, as 41 percent of non-heterosexuals currently smoked, and 28 percent of heterosexuals were current smokers. Never smokers comprised 37 percent of the non-heterosexuals and 51 percent of heterosexual, and former smokers represented 22 percent and 21 percent of non-heterosexuals and heterosexuals, respectively. The average age was 36.8 years for non-heterosexuals and 39.3 for heterosexuals and the age distribution was approximately normal based on probability plots of normality after weighting. This difference of mean age was statistically significant using a t-test ( $p = 0.02$ ). Non-heterosexuals rated their current health as excellent, very good, good, fair, and poor, 8, 40, 39, 9, and 4 percent of the time, respectively, in this sample. Heterosexuals rated their current health as excellent,

very good, good, fair, and poor, 15, 38, 34, 12, and 2 percent of the time, respectively, in this sample.

#### 4.2 Health indicator comparisons

After categorizing the individual biomarker values into clinically significant values and assigning a value of 1 for each significant value, a test of homogeneity between those observations falling into either clinically significant values, or non-significant values revealed that there was no departure from homogeneity across sexual identity categories at a statistical significance level of 0.05. Table 2 displays results of percentages of either sexual identity falling into the high risk group for each individual health index. Although no result was statistically significant at  $\alpha = 0.05$ , HDL cholesterol (62 percent of heterosexual and 71 percent of non-heterosexuals) was approaching statistical significance at  $p = 0.07$ . Glycohemoglobin was also somewhat different at  $p = 0.13$ , with 4.4 percent of heterosexuals and 8.0 percent of non-heterosexuals having high values.

There were standard errors for prevalence estimates (after weighting) among the non-heterosexuals that were greater than 30% of the actual estimate for glycohemoglobin, systolic blood pressure, and diastolic blood pressure. Prevalence estimates for glycohemoglobin, systolic blood pressure, and diastolic blood pressure should be viewed with caution for non-heterosexuals.

In the logistic regression models of clinically significant values against an exposure of sexual identity (correcting for pertinent risk factors), the only statistically significant association observed was for glycohemoglobin (glycated hemoglobin) (OR = 2.432, 95% CI 1.057, 5.596). More specifically, among the participants of the 2001-2004 cycles of NHANES IV (continuous), non-heterosexuals had 2.432 times the odds of

having clinically significant values of glycohemoglobin as compared to heterosexuals. These results are compatible with an odds ratio ranging from 1.057 to 5.596. HDL cholesterol did approach statistical significance (OR = 1.483 95% CI 0.931, 2.397) (see table 4). Table 4 contains odds ratios and 95% confidence intervals for having a clinically significant biomarker value in non-heterosexuals as compared to heterosexuals for each individual biomarker index of health.

### *4.3 Allostatic load*

After summing the individual indices, a logistic regression model was created. Allostatic load score of greater than 0 was considered the outcome of interest, with 0 being the comparison group. All data must have had responses for all biomarker variables of interest. Any observation that did not meet this criterion was excluded from the analysis of allostatic load. This reduced the sample to 2593 participants with 94 identifying as non-heterosexual. PIR, education, race and age were included in the model. Illegal drug use, more than moderate consumption of alcohol, current health status and tobacco use were tested in the model as potential confounders. Current health status influenced the beta coefficient of sexual identity (modeled against allostatic load score) by more than 10 percent and thus was included in the model. Current health status may affect allostatic load, and is different from cumulative effects on health by chronic stress and thusly is controlled for in the model. The odds ratio of having a score of 1 or greater (as compared to having a score of 0) in non-heterosexuals as compared to heterosexuals was 1.202 in the NHANES continuous module, 2001-2004. These data are consistent with a 95% confidence interval from 0.663 to 2.181.

#### *4.4 Propensity score matching*

All demographic variables and covariates were used to predict the probability of exposure (to homosexuality). After propensity score matching, there were 186 observations, of which 93 were non-heterosexuals. Table 3 displays the balance of demographics across sexual identity categories. All groups were balanced, and showed no heterogeneity across sexual identity (i.e., all chi-square statistics resultant from tests of homogeneity were greater than 0.05). Mean age was approximately 36 in both identity categories (for t-test of difference of means  $p = 0.852$ ). Education levels ( $p = 0.7596$ ), drinking above moderation ( $p = 0.8464$ ), ever used street drugs ( $p = 0.5039$ ), and smoking status ( $p = 0.9925$ ) were comparable after propensity score matching.

Logistic regression was performed by modeling sexual identity against an outcome of high risk values in a given biomarker of health. Table 4 contains the values of all odds ratios of having a high index value among non-heterosexuals as compared to heterosexuals and 95% CI's for each biomarker index after propensity score matching. Of particular interest is the increase of OR in HDL cholesterol to OR = 2.865 (95% CI 1.342, 6.117), whereas the lower limit of the 95% CI was 0.931 before propensity score matching. Also, glycohemoglobin could not be compared after propensity score matching as this excluded all heterosexuals with a high value of glycohemoglobin.

Odds ratio of having an allostatic load score of greater than 0 remained similar to the odds ratio of the non-propensity score matched data (OR = 1.299 95% CI (0.638, 2.646) (see table 4).

## CHAPTER 5

### DISCUSSION

Heterosexuals and non-heterosexuals were shown to be different in terms of alcohol drinking, illegal drug use, tobacco use and education. Our study is among the first to exam behavioral risk factors between heterosexuals and non-heterosexuals using nationally representative data. However, we did not find statistically significant differences in individual biomarkers indices of stress with the exception of glycohemoglobin. When comparing allostatic load as a summary indicator of stress induced biologic dysregulation, we did not observe an association between allostatic load and sexual identity. After propensity score matching results remained unchanged except that HDL cholesterol was positively associated with having sexual identity of non-heterosexual.

Non-heterosexuals and heterosexuals are shown to be dissimilar in regard to education, mean age, drinking behavior, illegal drug use, and smoking habits, echoing previous studies of health behaviors (Ungvarski & Grossman, 1999). Over 40% of non-heterosexuals smoke and 45% of non-heterosexuals have tried illegal street drugs in this sample. 64% drink above moderation when they drink, according to these data. These behaviors all have significant health implications and suggest that the health experience of non-heterosexuals will be different than that of heterosexuals should this be an accurate portrayal of the non-heterosexual experience.

Substance abuse has been a topic of concern for health officials for some time, and these data suggest that gay men, lesbians, and bisexuals may need targeted

interventions. Although several studies confirm that substance abuse is more prevalent in the gay community, prevalence estimates of drug, alcohol, and tobacco use are extremely high and may be due to small sample size of homosexual and bisexual participants. Also, the weighting scheme is designed for the purposes of making NHANES continuous modules nationally representative given that it over-samples geographic areas known to be high in racial minority populations and those people over 60 years of age. By weighting, this may create a population that is representative overall, but in terms of those participants that identify as gay men, lesbians, or bisexuals, the application of weighting may be creating a population that isn't reflective of the actual non-heterosexual population. Also, some cell sizes were small when categorizing participants according to their demographics which can affect the accuracy of these weighted measures. This is especially the case for PIR, in which the 3.0-3.99 and 4.0-4.99 categories for gay men, lesbians and bisexuals had only 6 and 8 participants, respectively. PIR, however, was still similar among non-heterosexuals and heterosexuals, but for non-heterosexuals the standard error was half of the actual estimate. Similarly, the standard error of the highest age category for non-heterosexuals was also about half of the estimate. This calls into question the prevalence estimates of non-heterosexuals in the United States for these strata of PIR. After propensity score matching, matched heterosexuals had similar standard errors to non-heterosexuals for the specific categories discussed, but these matched data are not being used as prevalence estimates for the population at large.

PIR (a proxy measure of socioeconomic position) has also demonstrated a negative relationship to the probability of having high allostatic load (Seeman, Merkin, et

al., 2008). Based on this analysis, PIR categories are statistically homogenous by sexual identity. PIR has been among the strongest predictors of allostatic load in previous studies, but should not be a confounding factor in this study, as it is both controlled for in the logistic regression model (of sexual identity as it predicts probability of being in a high allostatic load category) and was homogenous across sexual identity categories.

Education (also a proxy of socioeconomic position) has been shown to be negatively related to probability of having a high allostatic load score (Seeman, Merkin, et al., 2008). These data show that those people identifying as homosexual or bisexual are more likely to have additional education beyond the high school level. It would suggest that this may be somewhat protective based on previous literature on allostatic load. This may not be the case, however, as non-heterosexuals may not have the same family obligations as their heterosexual counterparts. For example, gay men and lesbians may not be as likely to have children and may not identify as strongly with the idea of a nuclear family. Social ties may be fostered in places where the traditional family structure is not emphasized, such as college. If that is the case, gay men's and lesbians' attainment of higher educational status may not be reflective of higher socioeconomic position, as they may not be as likely to have dependent children that would prevent them from achieving education beyond high school.

Race is also similar by sexual identity categories. Race has been shown to predict allostatic load. As with PIR, race has been controlled for in the allostatic load model.

Each individual biomarker of health was dichotomized by guidelines set forth in Teresa Seeman and colleague's (2008) study of allostatic load using NHANES III data.

The prevalence of clinically significant factors, for the most part, are similar across health indices. Odds of having high glycohemoglobin levels after correcting for relevant factors, however, are greater among non-heterosexuals than heterosexuals. Streptoe and Marmot (2003) demonstrated that glycohemoglobin levels were positively associated with psychosocial adversity. Interestingly enough, they also found no association between CRP and psychosocial adversity in their study. This is evidence that higher prevalence of clinically significant values of glycohemoglobin may be induced by the stress that non-heterosexuals experience. Further, CRP may not be indicative of stress induced health deterioration.

The main hypothesis of this study was that odds of having a high allostatic load is higher in those people not identifying as non-heterosexual. The odds of having allostatic load score of 1+ in non-heterosexuals was 1.202 times the odds of having allostatic load scores of 1+ in non-heterosexual among those NHANES participants aged 20-60 responding to the sexual behavior module with data for all relevant factors in 2001-2004. The 95% confidence interval was 0.663 to 2.181. The power of this analysis (based on the total sample size of 2687, exposed to unexposed ratio of 27 (2593 heterosexual / 94 non-heterosexuals, and a 66% prevalence of allostatic load 1+ in heterosexuals) exceeds 80%. If there was an effect of sexual identity on allostatic load, it is likely that it would have been observed. However, the analysis of complex survey data involves a high degree of variance and may affect the width of confidence intervals. The power estimates for a propensity score matched data set in which there are 186 observations are quite different. To achieve sufficient power (0.80) to detect an odds ratio of 1.3

(assuming 1 to 1 match for heterosexual and non-heterosexuals), there needed to be a sample size of 900. The resultant 95% confidence interval after matching was 0.638 to 2.646. After propensity score matching, if there was sufficient sample size in the non-heterosexual category to have a matched sample with  $n = 900$ , a more statistically sound conclusion could be drawn about allostatic load as it relates to stress associated with identifying as non-heterosexual. Analysis of homogeneity among demographics shows that the two sexual identity groups are more comparable as a result of propensity score matching; however, this comes at the expense of power.

There are several possible explanations for the lack of significant findings when comparing allostatic load score across sexual identities. The most obvious is that there may not be a difference in the stress experience between heterosexuals and homosexuals. Assuming that is not the case, as the literature presented here on gay, lesbian, and bisexual health suggests that it is not, there are other explanations to account for lack of statistically significant findings regarding allostatic load. First, the sample size of non-heterosexuals is very small ( $n = 167$  before propensity score matching,  $n=94$  for allostatic load analysis,  $n = 93$  after propensity score matching). If the effect of being non-heterosexual on the odds of having a high allostatic load is relatively small (OR below 1.5), there may not be adequate power to detect such a small difference. Also, not all people that are homosexual/bisexual will identify as such. The people who do may be systematically different in risk behaviors as well as allostatic load score from those identifying as non-heterosexual. Openly identifying oneself as non-heterosexual may suggest better adaptation to one's environment (compared to those who conceal their

sexuality) and those people openly identifying themselves as non-heterosexual may experience less stress associated with their sexual identity. Level of concealment may also be a significant contributor to the stress experience, but is not measured in NHANES. Additionally, biomarkers used for this analysis are well suited to detecting differences in other minority populations, but may not be well suited for use in this community. It may be necessary to explore other biological markers to compile an allostatic load score for sexual identity-based discrimination, or to only analyze certain biomarkers as this definition allostatic load may not reflect the health related stress experience of this population. Lastly, other extraneous variables may also exist that may confound the relationship between sexual identity and allostatic load that are not measured here.

After propensity score matching, despite seeing no difference in odds of having an allostatic load score, two interesting results did come out of this study in regard to individual biomarker indices. First, glycohemoglobin levels were even less homogenous than non-propensity score matched data. Second, clinically low HDL cholesterol values were not homogenous across sexual identity categories. In the non-propensity score matched data, there was an increased odds of having a high glycohemoglobin score among non-heterosexuals. When propensity score matched, there were no individuals with clinically significant values of glycohemoglobin among heterosexuals while 5.9 percent of non-heterosexuals had high glycohemoglobin levels.

Clinically low HDL cholesterol was not evenly experienced by heterosexuals and non-heterosexuals after propensity score matching. Although there were quite a number

of people with low levels of HDL cholesterol before propensity score matching in both sexual identity categories, after excluding the observations without values for variables needed for propensity score matching and performing matching there was a sharp decrease in the prevalence of low HDL cholesterol levels in both sexual identity groups. The distribution of clinically significant values of HDL cholesterol was not uniform across sexual identity categories after propensity score matching. There is little recent research that implicates low levels of HDL cholesterol with stress, but Wattoo, Memon, Memon, Wattoo, Tirmizi, and Iqbal (2008), found in a cohort of Pakistani college teachers and housewives that the stress experience of housewives was greater than that of college teachers and that housewives had lower HDL cholesterol levels than college teachers. While not directly claiming stress affects HDL cholesterol level, their research does support the idea that stress and HDL cholesterol are inversely related. In animal studies (using rats), after inducing “severe psychic trauma” HDL cholesterol levels decreased and were persistently low for six weeks following the stress occurred (Tsikunov, Klyueva, Kusov, Vinogradova, Klimenko, & Denisenko, 2006). Based on these two studies of HDL cholesterol and stress, a corollary may be drawn between the stress incumbent upon not identifying as heterosexual and decreased HDL cholesterol levels.

Although glycohemoglobin and HDL cholesterol are affected by identifying as non-heterosexual in this study, all other biomarkers used in this study were not affected. Gay men, lesbians, and bisexuals were younger on average than their heterosexual counterparts and only those individuals 20-60 years of age participated in the sexual

behavior module. It could be the case that non-heterosexuals have not had time to become chronically dysregulated in the other indices of health. It may take more time to become dysregulated in other biomarkers measured here, whereas glycohemoglobin and HDL cholesterol may be more proximal indicators of the effects of stress upon health. More proximal biomarker measures (in addition to glycohemoglobin and HDL cholesterol) may be necessary then to capture the stress-induced biologic dysregulation experience of gay men, bisexuals, and lesbians given the age profile of the respondents in this study.

It is necessary to quantify how values of each biomarker are related to stress, or if there is indeed a relationship at all. A modest amount of research has been conducted assessing allostatic load and minority stress, but further quantitative work on individual biomarkers and allostatic load models may be necessary. Also, the stress experience of gay men, lesbians, and bisexuals may be different than the stress experiences in other minority populations. It may be necessary to use other biomarkers to measure the biologic dysregulation in non-heterosexuals. The applicability of allostatic load (as it is defined here) may need to be assessed in terms of the specific minority population in which the stress-induced biologic dysregulation is being measured.

Although the primary hypothesis of this study is not supported by these data both before and after propensity score matching, there is evidence that the stress of identifying as non-heterosexual may be affecting the health of those that identify as non-heterosexual. Many groups intervene on behalf of homosexuals and bisexuals. If, as this study suggests, there are biological consequences of stress associated with being a gay

man, lesbian, or bisexual, it is necessary to evaluate the extent to which these groups reach the people who need their services and to then evaluate the efficacy of these services. This will provide a more complete picture of the health needs of non-heterosexuals and the pathway through which stress associated with identifying as non-heterosexual affects health.

## CHAPTER 6

### LIMITATIONS

There are several issues that may limit the ability of this body of research to address the objectives set forth. The most pressing among them is that the gay community may not be adequately sampled in NHANES. As is the case with other minority groups, this population would need to be over sampled to obtain an adequate representation of homosexuals and bisexuals. This could be impractical to implement as the NHANES sampling scheme selects by county and gay men and lesbians may not be clustered geographically in the population. Theoretically, this can limit the generalizability of the study. The measures of effect may not in this instance be representative of the general United States non-heterosexual population.

Secondly, there are issues of social desirability involved. This may influence a prevarication bias such that some people will not identify themselves as “gay” or having sex with someone of the same gender. If those misrepresenting themselves as heterosexual have similar clinical profiles as those identifying themselves as homosexual/bisexual and homosexuals/bisexuals have a dissimilar clinical profile to the heterosexual category, then the results will be biased toward the null. This is assuming there is a positive association between exposure and outcome. Also, there could be another information bias in that certain races/ethnicities may be less likely to identify themselves as “not heterosexual” than other ethnicities. This may result in differential misclassification of exposure. Assuming that these groups that are not heterosexual (but identify as heterosexual) have different clinical profiles from those that are truly

## CHAPTER 6

### LIMITATIONS

There are several issues that may limit the ability of this body of research to address the objectives set forth. The most pressing among them is that the gay community may not be adequately sampled in NHANES. As is the case with other minority groups, this population would need to be over sampled to obtain an adequate representation of homosexuals and bisexuals. This could be impractical to implement as the NHANES sampling scheme selects by county and gay men and lesbians may not be clustered geographically in the population. Theoretically, this can limit the generalizability of the study. The measures of effect may not in this instance be representative of the general United States non-heterosexual population.

Secondly, there are issues of social desirability involved. This may influence a prevarication bias such that some people will not identify themselves as “gay” or having sex with someone of the same gender. If those misrepresenting themselves as heterosexual have similar clinical profiles as those identifying themselves as homosexual/bisexual and homosexuals/bisexuals have a dissimilar clinical profile to the heterosexual category, then the results will be biased toward the null. This is assuming there is a positive association between exposure and outcome. Also, there could be another information bias in that certain races/ethnicities may be less likely to identify themselves as “not heterosexual” than other ethnicities. This may result in differential misclassification of exposure. Assuming that these groups that are not heterosexual (but identify as heterosexual) have different clinical profiles from those that are truly

heterosexual, this differential misclassification will decrease the measure of effect if the allostatic load is higher in non-heterosexuals. It could also decrease the measure of effect if the allostatic load is higher in truly heterosexual people for the individual logistic regression models of individual biomarkers. This will also cause artificial decrease or increase in odds of having an allostatic load score for those who identify as heterosexual if they are indeed not heterosexual. There will be a decrease or increase in allostatic load score for the heterosexual category as well, although due to the large number of heterosexual people, this will not be as drastic as it will be for the non-heterosexual category.

It could also be the case that those gay people who do not identify themselves as such may be systematically different in their risk of having high allostatic load scores / biologic dysregulation. If this is the case, it will induce a bias in which those that are inaccurately identifying themselves will bias the allostatic load score toward the direction (either higher or lower than the theoretic true value including all gay people) that they, on average, experience. This will bias the measure of association negatively or positively away from the null.

NHANES does not measure level of concealment of sexual identity.

Concealment can increase the stress experience of non-heterosexuals. Concealment should ideally be controlled for in modeling the odds of having an allostatic load score among non-heterosexuals as compared to heterosexuals.

Clinical measures may pose some measurement problems. For instance, BMI may be an inaccurate measure of obesity in those that are very muscular, or very short. If

any one group is more likely to be muscular or short, the measure of association for that biomarker will artificially inflate the measure of effect, depending on the nature of the association in the logistic regression of that individual factor. This will also artificially increase the allostatic load coefficients for the groups that include more muscular or short people and can increase the odds ratio of having an allostatic load score if the non-heterosexuals are systematically shorter or more muscular. Because measurement was standardized for all laboratory measures, there is no real concern of differential misclassification. In general, if there is misclassification, it will be non-differential and thusly tend to draw the measure of effect for the individual biomarkers toward the null value. The allostatic load scores should be uniformly biased from their true measure across all categories of sexual identity.

Smoking, alcohol use and illegal drug use variables were used for propensity score matching. It is shown that these are covariates of homosexuality and bisexuality (Ungvarski & Grossman, 1999; & Gay and Lesbian Medical Association, N.D.). However, it may not be socially desirable to identify as an illegal drug user, smoker or heavy alcohol user. Given that these exposures are more likely to be seen in these populations as opposed to heterosexuals, this can cause misclassification. Propensity score matching was implemented to reduce residual bias by matching on these (and other) covariates of the exposure of interest. If the prevalence of these predictors of exposure is underrepresented in the exposed (non-heterosexuals) due to prevarication then there is still residual confounding after propensity score matching as propensity estimates (of being non-heterosexual) will be inaccurate.

If the risk factors associated with the exposure chosen to be propensity score matched upon do not adequately reflect the counterfactual framework (and thus control for residual confounding), those measures of effect (odds ratios) as well as the allostatic load scores will not be reflective of the theoretic true values. This will depend, of course, on the nature of the confounders (risk factors) as to how these values are affected.

Also, if there is differential misclassification of the exposure, propensity score matching may cause the measures of effect to be biased toward the null and the allostatic load scores to be artificially similar among all sexual identity groups. Those people that incorrectly identify as heterosexual may still have similar clinical profiles to those that are correctly identify as not heterosexual. Secondly, they will likely have similar attributes for propensity score matching and more likely be chosen as unexposed controls. If this happens, propensity score matching will prove to be of less utility than intended and may show a decrease in measures of effect and outcome measures than when propensity score matching is not applied.

Lastly, this research project is based on the assumption that non-heterosexuals have increased stress as compared to heterosexuals. NHANES does not directly measure stress, although allostatic load score (as it is functionally defined here) is a proxy measure of stress induced biologic dysregulation. Since level of stress due to sexual orientation cannot be accurately quantified, it cannot be definitively stated that stress is the cause of dysregulation of allostatic load score or an individual biomarker level despite evidence that non-heterosexuals inherently have higher levels of stress as shown by the reviewed research of gay, lesbian and bisexual health.

## REFERENCES / BIBLIOGRAPHY

- Allsworth, J. E., Weitzen, S., and Boardman, L. A. (2005). Early age at menarche and allostatic load: data from the Third National Health and Nutritional Examination Survey. *Ann Epidemiology*, 15: 438-444.
- Centers for Disease Control and Prevention. (2007). Defining Overweight and Obesity. Retrieved April 1, 2008, from <http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm>.
- Centers for Disease Control and Prevention. (2007). High Blood Cholesterol Prevention. Retrieved April 1, 2008, from <http://www.cdc.gov/cholesterol/prevention.htm>.
- Centers for Disease Control and Prevention. (2007). Preventing and controlling high blood pressure. Retrieved April 1, 2008, from <http://www.cdc.gov/bloodpressure/prevention.htm>.
- Chu, K.C., Miller, B.A., Springfield, S.A. (2007). Measures of racial/ethnic health disparities in cancer mortality rates and the influence of socioeconomic status. *Journal of the National Medical Association*. 99: 1092-100, 1102-4.
- Cochran, S. D., & Mays, V. M. (2007). Physical health complaints among lesbian, gay men, and bisexual and homosexually experienced heterosexual individuals: results from the California quality of life survey. *American Journal of Public Health*, 97(11): 2048-2055.
- Cole, S. W., Kemeny, M. E., Taylor, S. E., & Visscher, B. R. (1996). Elevated physical health risk among gay men who conceal their homosexual identity. *Health Psychology*, 15(4): 243-251.

- Danesh, J., Collins, R., Appleby, P., Peto, R. (1998). Association of fibrinogen, c-reactive protein, albumin, or leukocyte count with coronary heart disease. *JAMA*. 279:1477-1482.
- Diamant, A. L., & Wold, C. (2003). Sexual orientation and variation in physical and mental health status among women. *Journal of Women's Health*, 12(1): 41-49.
- Gay and Lesbian Medical Association. (n.d.). Top Ten Issues to Discuss with Your Healthcare Provider. Retrieved January 2, 2008, from <http://www.glma.org/index.cfm?fuseaction=Page.viewPage&pageID=586>.
- Geronimus, A.T., Hicken, M., Keene, D., Bound, J. (2006). "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *AJPH*. 96(5): 826-33.
- Heck, J. E., & Jacobson, J. S. (2006). Asthma diagnosis among individuals in same-sex relationships. *Journal of Asthma*, 43: 579-584.
- Huebner, D. M., & Davis, M. C. (2005). Gay and bisexual men who disclose their sexual orientations in the workplace have higher workday levels of salivary cortisol and negative affect. *Ann. Behav. Med.*, 30(3): 260-267.
- Huebner, D. M., & Davis, M. C. (2007). Perceived antigay discrimination and physical health outcomes. *Health Psychology*, 26(5): 627-634.
- King, M., & Nazareth, I. (2006). The health of people classified as lesbian, gay and bisexual attending family practitioners of London: a controlled study. *BMC Public Health*, 6:127. Retrieved December 8, 2007 from <http://www.biomedcentral.com/1471-2458/6/127>.

- Krieger, N., & Sidney, S. (1997). Prevalence and health implication of anti-gay discrimination: A study of black and white women and men in the cardia cohort. *International Journal of Health Services*, 27(1): 157-76.
- Lock, J., & Steiner, H. (1999). Gay, lesbian, and bisexual youth risks for emotional, physical, and social problems: results from a community-based survey. *J. Am. Acad Child Adolesc. Psychiatry*, 38(3): 297-304.
- Mayo Clinic Staff. (2007). Tachycardia. Retrieved April 1, 2008, from <http://www.mayoclinic.com/health/tachycardia/DS00929/DSECTION=4>.
- McEwen, B. S. (2001). From molecules to mind. Stress, individual differences, and the social environment. *Ann. N. Y. Acad. Sci.*, 935: 42-9.
- National Diabetes Information Clearinghouse. (2007). Am I at risk for type 2 diabetes? Retrieved April 1, 2008, from <http://diabetes.niddk.nih.gov/dm/pubs/riskfortype2/risk.pdf>.
- Oakes, J. M., & Johnson, P. J. (2006). Propensity score matching for social epidemiology. In J. M. Oakes & J. S. Kaufman (Eds.), *Methods in Social Epidemiology* (pp. 370-392). San Francisco: John Wiley & Sons, Inc.
- Parsons, L. S. (2001). Reducing bias in a propensity score matched-pair sample using greed matching techniques. *SUGI 26*: 214-26.
- Perez-Benitez, C. I., O'Brien, W. H., Carels, R. A., Gordon, A. K., & Chiros, C. E. (2007). Cardiovascular correlates of disclosing homosexual orientation. *Stress and Health*, 23: 141-152.

- Seeman, T. E., Crimmins, E., Huang M., Singer, B., Bucur, A., Gruenewald, T., Berkman, L. F. and Reuben, D. B. (2004). Cumulative biological risk and socioeconomic differences in mortality: MacArthur studies of successful aging. *Social Science & Medicine*, 58(2004): 1985-1997.
- Seeman, T., Merkin, S.S, Crimmins, E., Koretz, B., Charette, S., Karlamangla, A. (2008). Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANESIII (1988-1994). *Social Science & Medicine*. 66(2008): 72-87.
- Seeman, T. E., Singer, B. H., Ruff, C. D., Love, G. D., & Levy-Storms L. (2002). Social relationships, gender, and allostatic load across two age cohorts. *Psychosomatic Medicine*, 64: 395-406.
- Skidmore, W.C., Linsenmeier, J.A.W., Bailey, J.M. (2006). Gender nonconformity and psychological distress in lesbians and gay men. *Archives of Sexual Behavior*, 35 (6): 685-697.
- Streptoe, A., Marmot, M. (2003). Burden of psychosocial adversity and vulnerability in middle age: association with biobehavioral risk factors and quality of life. *Psychosomatic Medicine*. 65 (6): 1029 – 1037.
- Tsikunov, S. G., Klyueva, N. N., Kusov, A. G., Vinogradova, T. V., Klimenko, V. M., Denisenko, A. D. (2006). Changes in the lipid composition of blood plasma and liver in rats induced by severe psychic trauma. *Bull Exp Biol Med*. 141(5): 636-8.
- United States Department of Agriculture and United States Department of Health and Human Services. (2005). Chapter 9 - Alcoholic Beverages. In: *Dietary Guidelines*

for Americans. (pp. 43-46) Washington, DC: US Government Printing Office;  
2005. p.43–46.

Ungvarski, P. J., & Grossman, A. H. (1999). Health problems of gay and bisexual men.  
*Nursing Clinics of North America*, 34(2): 313-331.

Wattoo, F. H., Memon, M. S., Memon, A. N., Wattoo, M. H. Tirmizi, S. A., Iqbal, J.  
(2008). Estimation and correlation of stress and cholesterol level in college  
teachers and housewives in Hyderabad-Pakistan. *J Pak Med Assoc.* 58(1): 15-8.

## APPENDIX

**Table 1. Key Demographic Characteristics across Sexual Identity, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module. ¶**

	Gay / Lesbian & Bisexual	Heterosexual	P-value **
<b>Mean age, yrs.</b>	36.8	39.3	0.02
<b>Age category</b>			
20-29	27.4	24.5	
30-39	31.2	25.3	
40-49	26.6	28.6	
50-60	14.8	21.6	0.28
<b>Gender (%female)</b>	50.2	50.0	0.97
<b>Education, yrs. (%)</b>			
Less than HS	11.5	14.8	
Completed HS	18.9	25.6	
More than HS	69.6	59.6	0.03
<b>Race (%)</b>			
White/Non-Hispanic	75.9	70.9	
Black/Non-Hispanic	6.1	8.1	
Mexican American	8.3	11.5	
Other	9.7	9.5	0.37
<b>Poverty income ratio (USD) (%)*</b>			
<1	12.4	12.5	
1.0-1.99	19.1	18.0	
2.0-2.99	21.0	15.1	
3.0-3.99	12.6	14.8	
4.0-4.99	8.1	13.3	
5.0+	26.7	26.3	0.51
<b>Drinking above moderation (%)</b>	64.6	55.1	0.05
<b>Ever used street drugs (%)</b>	45.0	21.2	<0.01
<b>Smoking status (%)</b>			
Current smoker	41.1	28.1	
Former smoker	21.9	20.8	
Never smoked	37.0	51.0	<0.01
<b>Self-Rated Current Health Status (%)</b>			
Excellent	8.3	15.1	
Very Good	39.5	37.6	
Good	38.9	33.5	
Fair	8.9	11.6	
Poor	4.4	2.3	0.07

¶. weighted percentages

\* defined as ratio of income to poverty threshold

\*\* p-value from chi-square test of homogeneity, except for mean age where independent t-test is used.

**Table 2: Percent of Participants with Clinically Significant Values\*\* of Individual Indices by Sexual Identity, United States, 2001-2004, NHANES Continuous Module.¶**

	Sexual Identity		p-value *
	Heterosexual	Non-heterosexual	
C-reactive protein	34.3	40.8	0.23
Albumin	9.2	10.2	0.55
Glyco - Hemoglobin	4.4	8.0	0.13
Total Cholesterol	15.1	12.3	0.42
HDL Cholesterol	61.9	70.6	0.07
BMI	31.1	34.0	0.38
Systolic Blood Pressure	5.4	4.9	0.88
Diastolic Blood Pressure	4.2	4.2	0.98
Pulse	9.6	8.6	0.68

¶. weighted percentages

\* p-value corresponds to chi square analysis of homogeneity of having clinically significant values across sexual identity categories

\*\* clinically significant values are defined as: albumin <3.8 g/dL, CRP ≥ 0.3 mg/dL, total cholesterol ≥ 240 mg/dL, HDL cholesterol < 40 mg/dL, glyco-hemoglobin ≥ 6.4 %, resting heart rate ≥ 90 beats per minute, systolic blood pressure ≥ 140 mm Hg, and diastolic blood pressure ≥ 90 mm Hg, and BMI ≥ 30.

**Table 3. Key Demographic Characteristics across Sexual Identity after Propensity Score Matching, United States, 2001-2004, National Health and Nutrition Examination Survey Continuous Module. ¶**

	Gay / Lesbian & Bisexual	Heterosexual	P-value**
<b>Mean age, yrs.</b>	36.4	36.1	0.85
<b>Age category</b>			
20-29	26.35	29.8	
30-39	36.48	35.3	
40-49	25.2	27.8	
50-60	11.96	7.1	0.52
<b>Gender (%female)</b>	47.28	53.4	0.48
<b>Education, yrs. (%)</b>			
Less than High School	12.22	13.8	
Completed High School	16.45	20.4	
More than High School	71.33	65.8	0.76
<b>Race (%)</b>			
White/Non-Hispanic	75.8	75.5	
Black/Non-Hispanic	6.2	10.3	
Mexican American	7.7	5.9	
Other	10.25	8.4	0.62
<b>Poverty income ratio (USD) (%)*</b>			
<1	8.9	20.1	
1.0-1.99	18.27	18.4	
2.0-2.99	17.48	11.1	
3.0-3.99	11.99	4.6	
4.0-4.99	8.28	7.8	
5.0+	35.08	38	0.34
<b>Drinking above moderation (%)</b>	63.55	65	0.85
<b>Ever used street drugs (%)</b>	47.49	53.5	0.50
<b>Smoking status (%)</b>			
Current smoker	43.03	45.3	
Former smoker	12.7	17.8	
Never smoked	44.28	36.9	0.99
<b>Self-Rated Current Health Status (%)</b>			
Excellent	9.53	10.6	
Very Good	48.52	40.3	
Good	33.46	35.7	
Fair	5.58	12.2	
Poor	2.9	1.3	0.42

¶. weighted percentages

\* defined as ratio of income to poverty threshold

\*\* p-value from chi-square test of homogeneity, except for mean age where independent t-test is used.

**Table 4: Odds Ratio (and 95% CI) of Having a Clinically Significant Index Value among Non-Heterosexuals as compared to Heterosexuals With and Without Matching, United States, 2001-2004, NHANES Continuous Module in Specific Biologic Indices. ¶**

Dichotomous Biologic Index	Non-Propensity Score Matched		Propensity Score Matched	
	Odds Ratio (95% CI)		Odds Ratio (95% CI)	
diastolic blood pressure^^	1.170	(0.291, 4.705)	9.523	(0.430, 210.843)
systolic blood pressure^^	1.014	(0.286, 3.587)	6.138	(0.599, 62.925)
glycohemoglobin****	2.432	(1.057, 5.596)	--*	
c-reactive protein (CRP) ^	1.228	(0.785, 1.921)	0.620	(0.294, 1.308)
albumin**	1.157	(0.758, 1.766)	0.368	(0.061, 2.218)
HDL cholesterol***	1.493	(0.931, 2.397)	2.865	(1.342, 6.117)
total cholesterol***	0.689	(0.360, 1.318)	0.691	(0.305, 1.564)
Heart rate^^^	1.154	(0.638, 2.088)	1.553	(0.369, 6.527)
BMI*****	1.187	(0.866, 1.625)	1.245	(0.594, 2.609)

**Odds ratio of having an allostatic load score of 1+ ^^:**

1.202	(0.663, 2.181)	1.299	(0.638, 2.646)
-------	----------------	-------	----------------

¶, logistic regression used to model odds of having clinically significant values and odd of allostatic load score

\* there were no observations with high glycohemoglobin after propensity score matching among heterosexuals while 5.9% of non-heterosexuals had high glycohemoglobin levels

\*\*controls for smoking, total cholesterol level, SEP (proxied by PIR) and BMI

\*\*\*controls for BMI, SEP (proxied by PIR), ethnicity, smoking status, alcohol use, and gender

\*\*\*\*controls for BMI, age, race and total cholesterol

\*\*\*\*\*control for race and age

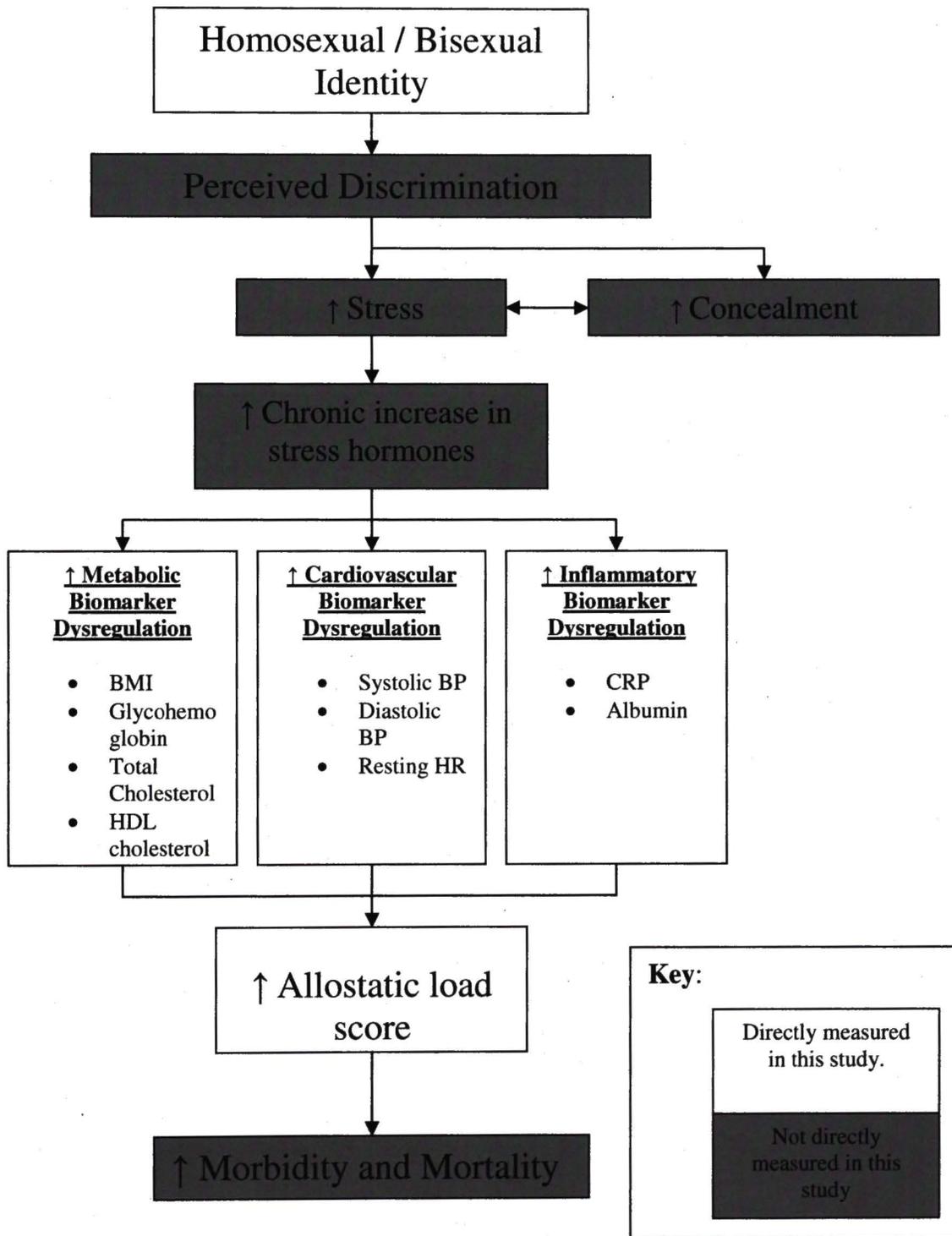
^ controls for BMI, smoking status, diabetes (proxied by glycohemoglobin values), smoking, and ethnicity

^^ controls for BMI, alcohol use, diabetes (proxied by glycohemoglobin values), smoking status and race

^^^ controls for age, race, and high blood pressure

^^^ controls for age, PIR, education and race

Figure1. Conceptual model relating sexual identity to allostatic load.



\*thesis code file 1 of 3 for power calculations;

```
/*-----  
*/  
/* Module name: macros/power.sas  
*/  
/* Function: Creates a power table for specified conditions  
*/  
/* given a range of measures of association and  
*/  
/* sample size.  
*/  
/*  
*/  
/* Input: Macro parms described below.  
*/  
/* Output: SAS dataset, GIF files for table and power chart.  
*/  
/*  
*/  
/* Parms: data - file name  
*/  
/* c - ratio of unexposed : exposed,  
*/  
/* or controls : cases  
*/  
/* p0 - prevalence of exposure in contols,  
*/  
/* or proportion of diseased in unexposed  
*/  
/* n - minimum sample size in range of  
sample size */  
/* n1 - maximum sample size in range of  
sample size */  
/* n2 - units in sample size range  
*/  
/* or - minimum OR in range of OR  
*/  
/* or1 - maximum OR in range of OR  
*/  
/* or2 - units in OR range  
*/  
/*  
*/  
/* Usage Notes:  
*/  
/*  
*/  
/*-----  
*/  
/* Maintenance History:  
*/  
/* Date Made By Description  
*/
```

```

/* -----
*/
/* Unknown F.F. Zhang Program created.
*/
/* 04/06/08 J.P. Adams Program adapted to macro format.
*/
/*-----
*/

```

```
options formdlim = '-' nomprint nonumber nodate;
```

```

%macro power (data = , /* file name
*/
                c = , /* ratio of unexposed : exposed,
*/
                /* or controls : cases
*/
                p0 = , /* prevalence of exposure in contols,
*/
                /* or proportion of diseased in
unexposed */
                n = , /* minimum sample size in range of sample
size */
                n1 = , /* maximum sample size in range of sample
size */
                n2 = , /* units in sample size range
*/
                or = , /* minimum OR in range of OR
*/
                or1 = , /* maximum OR in range of OR
*/
                or2 = /* units in OR range
*/
                );

```

```

%global data1;
%let data1 = &data;

```

```
Data &data;
```

```

c = &c;
p0 = &p0;
do n = &n to &n1 by &n2;
do or = &or to &or1 by &or2;

q0 = 1 - p0;
p1 = (p0*or) / (1 + p0 * (or - 1));
q1 = 1 - p1;
pbar = (p1 + c*p0) / (1 + c);
qbar = 1 - pbar;

zbeta = (sqrt (n * (p1 - p0)**2) -

```

```

        1.96 * sqrt (( 1 + 1 / c) * pbar * qbar)) / sqrt ((p1 * q1)
+ p0 * q0 / c);
        power = probnorm (zbeta);
        put power;
        output;
        end;
        end;

```

```

ods printer file = "C:\Documents and
Settings\64190\Desktop\&data" printer=gif;

```

```

proc print data = &data noobs;
    var p0 zbeta or n power;
run;

```

```

ods printer close;

```

```

%mend;

```

```

%macro power_plot;

```

```

ods printer file = "C:\Documents and
Settings\64190\Desktop\&data1.1" printer=gif;

```

```

proc gplot data = &data1;
    plot power*n = or; symbol line=1 width=5
interpol=join;
    title "Power curve, for work.&data1";
run;
quit;

```

```

ods printer close;

```

```

%mend;

```

```

%power (data = power,
        c      = 1,
        p0     = 0.5,
        n      = 0,
        n1     = 1000,
        n2     = 50,
        or     = 1,
        or1    = 2,
        or2    = 0.1
);

```

```

%power_plot;

```

```

/*****

```

```

/* Author: John Adams */
/* Output files: C:\Documents */
andSettings\e64190\Desktop\Thesis\complete */
/* C:\Documents and */
Settings\e64190\Desktop\Thesis\completea */
/* Title: Thesis code, part 2 of 3, data file code */
/*****/

```

```
libname thesis 'C:\Documents and Settings\e64190\Desktop\Thesis';
```

```
options fmtsearch = (thesis) nofmterr;
```

```
proc format library = thesis;
```

```

value agec 1 = '20-29'
           2 = '30-39'
           3 = '40-49'
           4 = '50-60'
           . = 'Missing';

```

```

value pir 1 = '5+'
          2 = '4.0-4.99'
          3 = '3.0-3.99'
          4 = '2.0-2.99'
          5 = '1.0-1.99'
          6 = '0.0-0.99'
          . = 'Missing';

```

```

value race 1 = 'White'
           2 = 'Mexican American'
           3 = 'Black'
           4 = 'Other'
           . = 'Missing';

```

```

value sxq 0 = 'Heterosexual'
          1 = 'Non-heterosexual';

```

```

value smoke 1 = 'Current smoker'
            2 = 'Former smoker'
            3 = 'Never smoker'
            . = 'Missing';

```

```
run;
```

```
options formdlm = '-' nodate;
```

```

libname dm xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\demo_b.xpt';
*contains seqn, dmdmart1, dmddeduc, indfminc, riagendr, ridageyr,
ridreth1;

```

```

data demo_b (keep = seqn WIMEC2YR SDMVPSU SDMVSTRA dmdmart1 dmddeduc
indfminc riagendr ridageyr ridreth1 INDFMPIR);

```

```
set dm.demo_b;
```

```
run;
```

```
libname smq_b xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\smq_b.xpt';  
*contains alq130;
```

```
data smq_b (keep = seqn smq020 smq040);  
set smq_b.smq_b;
```

```
run;
```

```
libname alq_b xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\alq_b.xpt';  
*contains alq130;
```

```
data alq_b (keep = seqn alq130);  
set alq_b.alq_b;
```

```
run;
```

```
data alcohol;  
set alq_b.alq_b;  
if alq120u = 1 and alq120q >=3 then it=1;
```

```
run;
```

```
proc freq data = alcohol;  
tables it;  
run;
```

```
libname duq_b xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\duq_b.xpt';  
*contains duq_b;
```

```
data duq_b (keep = seqn duq100);  
set duq_b.duq_b;
```

```
run;
```

```
libname hsq_b xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\hsq_b.xpt';  
*contains hsq_b;
```

```
data hsq_b (keep = seqn hsd010);  
set hsq_b.hsq_b;
```

```
run;
```

```
libname sx xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\sxq_b.xpt';  
*contains sxq292, sxq294 (sexual orientation);
```

```
data sxq_b (keep = seqn sxq292 sxq294);  
set sx.sxq_b;
```

```
run;
```

```
libname bpx xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2001\bpx_b.xpt';
```

```
    *contains bpxsy1, bpxsy2, bpxsy3 (systolic bp), bpxpls (pulse)
and bpxdi1, bpxdi2, bpxdi3 (diastolic bp);
```

```
data bpx_b (keep = seqn bpxsy1 bpxsy2 bpxsy3 bpxdi1 bpxdi2 bpxdi3
bpxpls);
```

```
    set bp.bpx_b;
run;
```

```
libname bx xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\bm_x_b.xpt';
    *contains BMXBMI (bmi);
```

```
data bm_x_b (keep = seqn bmxbmi);
    set bx.bm_x_b;
```

```
run;
```

```
libname l10 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\l10_b.xpt';
    *conatins lbxgh (glycohemoglobin);
```

```
data l10_b (keep = seqn lbxgh);
    set l10.l10_b;
```

```
run;
```

```
libname l11 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\l11_b.xpt';
    *conatins lbxcrp (crp);
```

```
data l11_b (keep = seqn lbxcrp);
    set l11.l11_b;
```

```
run;
```

```
libname l13 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\l13_b.xpt';
    *contains lbdhdl (hdl) lbxtc (total cholesterol);
```

```
data l13_b (keep = seqn lbdhdl lbxtc);
    set l13.l13_b;
```

```
run;
```

```
libname l40 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\l40_b.xpt';
    *contains LBDSR (creatinine) LBXSAL(albumin);
```

```
data l40_b (keep = seqn lbdscr lbxsal);
    set l40.l40_b;
```

```
run;
```

```
libname l06 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2001\l06_b.xpt';
    *contains lbdhcy (homocysteine);
```

```
data l06_b (keep = seqn lbdhcy);
    set l06.l06_b;
```

run;

data exam2001;

```
merge demo_b alq_b sxq_b (in=a) bpx_b bmx_b l10_b
      111_b 113_b 140_b 106_b smq_b duq_b hsq_b;
by seqn;
if a;
if lbdhcy =. then delete;
if LBDSCR =. then delete;
if LBXSAL =. then delete;
if lbdhd1 =. then delete;
if lbxtc =. then delete;
if lbxcrp =. then delete;
if lbxgh =. then delete;
if bmxbmi =. then delete;
if bpxsy1 =. then delete;
if bpxsy2 =. then delete;
if bpxsy3 =. then delete;
if bpxdi1 =. then delete;
if bpxdi2 =. then delete;
if bpxdi3 =. then delete;
if bpxpls =. then delete;
```

```
    bpxsy = (bpxsy1 + bpxsy2 + bpxsy3) / 3;
    bpxdi = (bpxdi1 + bpxdi2 + bpxdi3) / 3;
```

```
select;
when (sxq292= 1 or sxq294 = 1) sxq = 0;
when (sxq292= 2 or sxq294 = 2) sxq = 1;
when (sxq292= 3 or sxq294 = 3) sxq = 1;
otherwise                          sxq = .;
end;
```

run;

data exam2001a;

```
merge demo_b alq_b sxq_b (in=a) bpx_b bmx_b l10_b
      111_b 113_b 140_b 106_b smq_b duq_b hsq_b;
by seqn;
```

```
select;
when (sxq292= 1 or sxq294 = 1) sxq = 0;
when (sxq292= 2 or sxq294 = 2) sxq = 1;
when (sxq292= 3 or sxq294 = 3) sxq = 1;
otherwise                          sxq = .;
end;
```

run;

```

libname dm3 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\demo_c.xpt';
    *contains sqn, dmdmartl, dmdeduc, indfminc, riagendr,
ridageyr,ridreth1;

data demo_c (keep = seqn WTMEC2YR SDMVPSU SDMVSTRA dmdmartl dmdeduc
indfminc riagendr ridageyr ridreth1 INDFMPIR);
    set dm3.demo_c;
run;

libname alq_c xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\alq_c.xpt';
    *contains alq130;

data alq_c (keep = seqn alq130);
    set alq_c.alq_c;
run;

libname duq_c xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\duq_c.xpt';
    *contains duq_c;

data duq_c (keep = seqn duq100);
    set duq_c.duq_c;
run;

libname hsq_c xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\hsq_c.xpt';
    *contains hsq_c;

data hsq_c (keep = seqn hsd010);
    set hsq_c.hsq_c;
run;

libname smq_c xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\smq_c.xpt';
    *contains smq020 smq040;

data smq_c (keep = seqn smq020 smq040);
    set smq_c.smq_c;
run;

libname sx3 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\sxq_c.xpt';
    *contains sxq292, sxq294;

data sxq_c (keep = seqn sxq292 sxq294);
    set sx3.sxq_c;
run;

libname bp3 xport 'C:\Documents and
Settings\e64190\Desktop\Thesis\2003\bp3_c.xpt';

```

```
    *contains bpxsy1, bpxsy2, bpxsy3 (systolic bp), bpxpls (pulse),  
and bpxdi1, bpxdi2, bpxdi3 (diastolic bp);
```

```
data bpx_c (keep = seqn bpxpls bpxsy1 bpxsy2 bpxsy3 bpxdi1 bpxdi2  
bpxdi3);  
    set bp3.bpx_c;  
run;
```

```
libname bx3 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\bxm_c.xpt';  
    *contains BMXBMI (bmi);
```

```
data bmx_c (keep = seqn bmx bmi);  
    set bx3.bmx_c;  
run;
```

```
libname l103 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\l10_c.xpt';  
    *contains lbxgh (glycohemoglobin);
```

```
data l10_c (keep = seqn lbxgh);  
    set l103.l10_c;  
run;
```

```
libname l113 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\l11_c.xpt';  
    *contains lbd fbsi (fibrinogen, only in 01, not 03-04) lbxcrp  
(crp);
```

```
data l11_c (keep = seqn lbxcrp);  
    set l113.l11_c;  
run;
```

```
libname l133 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\l13_c.xpt';  
    *contains lbd hdd (hdl) LB DTCSI (total cholesterol);
```

```
data l13_c (keep = seqn lbxhdd lbxtc);  
    set l133.l13_c;  
run;
```

```
libname l403 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\l40_c.xpt';  
    *contains lbd strsi (triglycerides) LBxSCR (creatinine)  
LBXSAL(albumin);
```

```
data l40_c (keep = seqn lbd strsi lbxscr lbxsal);  
    set l403.l40_c;  
run;
```

```
libname l063 xport 'C:\Documents and  
Settings\64190\Desktop\Thesis\2003\l06mh_c.xpt';
```

```

        *contains lbdhcy (homocysteine);

data l06mh_c (keep = seqn lbxhcy);
    set l063.l06mh_c;
run;

data exam2003;
    merge demo_c sxq_c (in=a) alq_c bpx_c bmx_c l10_c
          l11_c l13_c l40_c l06mh_c smq_c duq_c hsq_c;
    by seqn;
    if a;
    if lbxhcy =. then delete;
    lbdhcy = lbxhcy;
    if lbdstrsi =. then delete;
    if LBxSCR =. then delete;
    lbdscr = lbxscr;
    if LBXSAL =. then delete;
    if lbxhdd =. then delete;
    lbdhdl = lbxhdd;
    if lbxtc =. then delete;
    if lbxcrp =. then delete;
    if lbxgh =. then delete;
    if bmxbmi =. then delete;
    if bpxsy1 =. then delete;
    if bpxsy2 =. then delete;
    if bpxsy3 =. then delete;
    if bpxdi1 =. then delete;
    if bpxdi2 =. then delete;
    if bpxdi3 =. then delete;
    if bpxpls =. then delete;

    bpxsy = (bpxsy1 + bpxsy2 + bpxsy3) / 3;
    bpxdi = (bpxdi1 + bpxdi2 + bpxdi3) / 3;

    select;
    when (sxq292= 1 or sxq294 = 1) sxq = 0;
    when (sxq292= 2 or sxq294 = 2) sxq = 1;
    when (sxq292= 3 or sxq294 = 3) sxq = 1;
    otherwise                      sxq = .;
    end;

run;

data exam2003a;
    merge demo_c sxq_c (in=a) alq_c bpx_c bmx_c l10_c
          l11_c l13_c l40_c l06mh_c smq_c duq_c hsq_c;

    by seqn;

    select;
    when (sxq292= 1 or sxq294 = 1) sxq = 0;
    when (sxq292= 2 or sxq294 = 2) sxq = 1;

```

```

when (sxq292= 3 or sxq294 = 3) sxq = 1;
otherwise                sxq = .;
end;

```

```
run;
```

```
data thesis.completea;
```

```
set exam2001a exam2003a;
```

```

bpxsy = (bpxsy1 + bpxsy2 + bpxsy3) / 3;
bpxdi = (bpxdi1 + bpxdi2 + bpxdi3) / 3;

```

```

if lbxsal < 3.8 then albumin = 1;
else                albumin = 0;

```

```

if lbxcrp >= 0.3 then crp = 1;
else                crp = 0;

```

```

if bmxbmi >= 30 then BMI = 1;
else                BMI = 0;

```

```

if lbxtc >= 240 then totchol = 1;
else                totchol = 0;

```

```

if lbdhdl < 40 then hdl = 1;
else                hdl = 0;

```

```

if lbxgh >= 6.4 then ghemo = 1;
else                ghemo = 0;

```

```

if bpxpls >= 90 then hr = 1;
else                hr = 0;

```

```

if bpxsy >= 140 then sbp = 1;
else                sbp = 0;

```

```

if bpxdi >= 90 then dpb = 1;
else                dpb = 0;

```

```

al = albumin + crp + BMI + totchol + hdl + ghemo + hr +
sbp + dpb;

```

```
if al >= 1 then allog1 = 1;
```

```
if 0 <= al < 1 then allog1 = 0;
```

```
mec4yr = 1/2 * WTMEC2YR;
```

```

if riagendr = 1 and alq130 > 2 then alcohol = 1;
else if riagendr = 1 and alq130 <= 2 then alcohol = 0;

```

```

if riagendr = 2 and alq130 > 1 then alcohol = 1;
else if riagendr = 2 and alq130 <= 1 then alcohol = 0;

```

```

if                                alq130 = . then alcohol = .;

if      0 < ridageyr < 30 then agec = 1;
else if 30 <= ridageyr < 40 then agec = 2;
else if 40 <= ridageyr < 50 then agec = 3;
else if 50 <= ridageyr      then agec = 4;
else if ridageyr = .        then agec = .;

if dmdeduc in (.,7,9) then dmdeduc = .;

if      0.0 < indfmpir < 1.0 then pir = 6;
else if 1.0 <= indfmpir < 2.0 then pir = 5;
else if 2.0 <= indfmpir < 3.0 then pir = 4;
else if 3.0 <= indfmpir < 4.0 then pir = 3;
else if 4.0 <= indfmpir < 5.0 then pir = 2;
else if 5.0 <= indfmpir      then pir = 1;
else if      indfmpir = .    then pir = .;

if      ridreth1 = 1 then race = 2;
else if ridreth1 = 3 then race = 1;
else if ridreth1 = 4 then race = 3;
else if (ridreth1 = 2 or
        ridreth1 = 5) then race = 4;
else      race = .;

if smq020 = 2                                then smoke = 3;
else if smq020 = 1 and smq040 = 3            then smoke = 2;
else if (smq020 = 1 and smq040 = 1) or
        (smq020 = 1 and smq040 = 2)          then smoke = 1;
else      smoke = .;

if (duq100 = 7) or (duq100 = 9) then duq100= .;

if      hsd010 in (1,2,3) then health = 1;
else if hsd010 in (4,5)  then health = 2;
else if hsd010 = .      then health = .;

format agec agec. pir pir. race race. sxq sxq. smoke smoke.;

run;

run;

data thesis.complete;

set exam2001 exam2003;

if lbxsal < 3.8 then albumin = 1;
else      albumin = 0;
if lbxcrp >= 0.3 then crp = 1;
else      crp = 0;
if bmxbmi >=30 then BMI = 1;

```

```

else BMI = 0;
if lbxtc >= 240 then totchol = 1;
else totchol = 0;
if lbdhdl < 40 then hdl = 1;
else hdl = 0;
if lbxgh >= 6.4 then ghemo = 1;
else ghemo = 0;
if bpxpls >= 90 then hr = 1;
else hr = 0;
if bpxsy >= 140 then sbp = 1;
else sbp = 0;
if bpxdi >= 90 then dpb = 1;
else dpb = 0;

if lbxsal < 4.0 then albumin1 = 1;
else albumin1 = 0;
if lbxcrp >= 0.47 then crp1 = 1;
else crp1 = 0;
if bmx bmi > 31.53 then BMI1 = 1;
else BMI1 = 0;
if lbxtc >= 227 then totchol1 = 1;
else totchol1 = 0;
if lbdhdl < 42 then hdl1 = 1;
else hdl1 = 0;
if lbxgh >= 5.5 then ghemo1 = 1;
else ghemo1 = 0;
if bpxpls >= 80 then hr1 = 1;
else hr1 = 0;
if bpxsy >= 125.33 then sbp1 = 1;
else sbp1 = 0;
if bpxdi >= 78.66 then dpb1 = 1;
else dpb1 = 0;
if lbdhcy > 9.11 then homo = 1;
else homo = 0;

al = albumin + crp + BMI + totchol + hdl + ghemo + hr +
sbp + dpb;
all = albumin1 + crp1 + BMI1 + totchol1 + hdl1 + ghemo1 + hr1 +
sbp1 + dpb1;

if al > 3 then allog = 1;
if 0 <= al <= 3 then allog = 0;

if al >= 1 then allog1 = 1;
if 0 <= al < 1 then allog1 = 0;

metabol = ghemo + totchol + hdl + bmi;

if metabol >= 1 then almeta = 1;
if metabol < 1 then almeta = 0;

inflamm = crp + albumin;

if inflamm >= 1 then alinfl = 1;

```

```

if inflamm < 1 then alinfl = 0;

cardio = sbp + dpb + hr;

if cardio >=1 then alcard = 1;
if cardio < 1 then alcard = 0;

mec4yr = 1/2 * WTMEC2YR;

if duq100 = . then delete;

if      riagendr = 1 and alq130 > 2 then alcohol = 1;
else if riagendr = 1 and alq130 <= 2 then alcohol = 0;

if      riagendr = 2 and alq130 > 1 then alcohol = 1;
else if riagendr = 2 and alq130 <= 1 then alcohol = 0;

if      alq130 = . then delete;

if      0 < ridageyr < 30 then agec=1;
else if 30 <= ridageyr < 40 then agec=2;
else if 40 <= ridageyr < 50 then agec=3;
else if 50 <= ridageyr      then agec=4;
else if ridageyr = .      then delete;

if dmdeduc in (.,7,9) then delete;

if      0.0 < indfmpir < 1.0 then pir = 6;
else if 1.0 <= indfmpir < 2.0 then pir = 5;
else if 2.0 <= indfmpir < 3.0 then pir = 4;
else if 3.0 <= indfmpir < 4.0 then pir = 3;
else if 4.0 <= indfmpir < 5.0 then pir = 2;
else if 5.0 <= indfmpir      then pir = 1;
else if indfmpir = .      then delete;

if      ridreth1 = 1 then race = 2;
else if ridreth1 = 3 then race = 1;
else if ridreth1 = 4 then race = 3;
else if (ridreth1 = 2 or
        ridreth1 = 5) then race = 4;
else delete;

if smq020 = 2      then smoke = 3;
else if smq020 = 1 and smq040 = 3      then smoke = 2;
else if (smq020 = 1 and smq040 = 1) or
        (smq020 = 1 and smq040 = 2)      then smoke = 1;
else delete;

if (duq100 = 7) or (duq100 = 9) then delete;

if      hsd010 in (1,2,3) then health = 1;
else if hsd010 in (4,5) then health = 2;
else if hsd010 = .      then delete;

```

```

if sxq = . then delete;

format agec agec. pir pir. race race. sxq sxq. smoke smoke.;

run;

/* ***** */
/* Perform a Logistic Regression and save*/
/* the propensity score data set */
/* STUDY.Propen for all patients in the */
/* observational study. */
/* Statistic Name = PROB */
/* Note: PARMLABEL is a SAS Verion 8.0 */
/* option. */
/* ***** */

/* ***** */
/* Greedy 5->1 Digit Matching Macro */
/* ***** */

options mstored sasstore = macros;
libname macros 'C:\Documents and
Settings\64190\Desktop\Thesis\macros';

%MACRO GREEDMTCH
(
    Lib, /* Library Name */
    Dataset, /* Data set of all */
            /* patients */
    depend, /* Dependent variable */
            /* that indicates */
            /* Case or Control; */
            /* Code 1 for Cases, */
            /* 0 for Controls */
    matches /* Output file of matched */
            /* pairs */
) / store source des = 'greedy match for psm';

/* Macro to sort the Cases and Controls dataset */
%MACRO SORTCC;

proc sort data=tcases out=&LIB..Scase;
    by prob;
run;

proc sort data=tctrl out=&LIB..Scontrol;
    by prob randnum;
run;

%MEND SORTCC;

/* Macro to Create the initial Case and Control Data Sets */

```

```

%MACRO INITCC(digits);

    data tcases (drop = cprob) tctrl (drop = aprob) ;

        set &LIB..&dataset. ;

        /* Create the data set of Controls*/
        if &depend. = 0 and prob ne . then do;
            cprob = Round(prob,&digits.);
            Cmatch = 0;
            Length RandNum 8;
            RandNum=ranuni(1234567);
            Label RandNum='Uniform Randomization
Score';

            output tctrl;
        end;

        /* Create the data set of Cases */
        else if &depend. = 1 and prob ne . then do;
            Cmatch = 0;
            aprob =Round(prob,&digits.);
            output tcases;
        end;
run;

%SORTCC;

%MEND INITCC;

/* Macro to Perform the Match */
%MACRO MATCH (MATCHED,DIGITS);
    data &lib..&matched. (drop = Cmatch randnum aprob cprob
matched);
                                start oldi curctrl

        /* select the cases data set */
        set &lib..SCase ;

        curob + 1;
        matchto = curob;
        if curob = 1 then do;
            start = 1;
            oldi = 1;
        end;

        /* select the controls data set */
        DO i = start to n;

            set &lib..Scontrol point= i nobs = n;

            if i gt n then goto startovr;
            if _Error_ = 1 then abort;
            curctrl = i;

```

```

/* output control if match found */
if aprob = cprob then do;
Cmatch = 1;
output &lib..&matched.;
matched = curctrl;
goto found;
end;

/* exit do loop if out of potential
matches */
else if cprob gt aprob then goto
nextcase;

startovr: if i gt n then goto nextcase;
END;
/* end of DO LOOP */

/* If no match was found, put pointer
back*/
nextcase: if Cmatch=0 then start = oldi;

/* If a match was found, output case and
increment pointer */
found: if Cmatch = 1 then do;
    oldi = matched + 1;
    start = matched + 1;
    set &lib..SCase point =
curob;
        output &lib..&matched.;
end;

retain oldi start;
    if _Error_=1 then _Error_=0;
run;

/* Get files of unmatched cases and */
/* controls. Note that in the example */
/* data, the patient identifiers are HID*/
/* (Hospital ID) and PATIENTN (Patient */
/* identifier. All cases have complete */
/* data for these two fields. Modify */
/* these fields with the appropriate */
/* patient identifier field(s) */

proc sort data=&lib..scase out=sumcase;
    by seqn;
run;

proc sort data=&lib..scontrol out=sumcontrol;
    by seqn;
run;

proc sort data=&lib..&matched. out=smatched (keep=seqn
matchto);

```

```

        by seqn;
run;

data tcases (drop=matchto);
    merge sumcase(in=a) smatched;
        by seqn;
    if a and matchto = . ;
    cmatch = 0;
    aprob =Round(prob,&digits.);
run;

data tctrl (drop=matchto);
    merge sumcontrol(in=a) smatched;
        by seqn;
    if a and matchto = . ;
    cmatch = 0;
    cprob = Round(prob,&digits.);
run;

%SORTCC

%MEND MATCH;

/* Note: This section can be          */
/* modified to try variations of the */
/* basic algorithm.                   */
/* Create file of cases and controls */
%INITCC(.00001);

/* Do a 5-digit match */
%MATCH(Match5,.0001);

/* Do a 4-digit match on remaining unmatched */
%MATCH(Match4,.001);

/* Do a 3-digit match on remaining unmatched */
%MATCH(Match3,.01);

/* Do a 2-digit match on remaining unmatched */
%MATCH(Match2,.1);

/* Do a 1-digit match on remaining unmatched */
%MATCH(Match1,.1);

/* Merge all the matches into one file */
/* The purpose of the marchto variable */
/* is to identify matched pairs for the*/
/* matched pair anlayses. matchto is */
/* initially assigned the observation */
/* number of the case. Since there */
/* would be duplicate numbers after the*/
/* individual files were merged, */
/* matchto is incremented by file. */
/* Note that if the controls file */

```

```

/* contains more than N=100,000 records*/
/* and/or there are more than 1,000 */
/* matches made at each match level, */
/* then the incrementation factor must */
/* be changed. */
data &lib..&matches.;
    set &lib..match5(in=a)
        &lib..match4(in=b)
        &lib..match3(in=c)
        &lib..match2(in=d)
        &lib..match1(in=e);

        if b then matchto=matchto + 100000;
        if c then matchto=matchto + 10000000;
        if d then matchto=matchto + 1000000000;
        if e then matchto=matchto + 1000000000000;

run;

/* Sort file -- Need sort for Univariate analysis in tables*/
proc sort data=&lib..&matches. out = &lib..S&matches.;
    by &depend.;
run;

%MEND GREEDMTCH;

```

```

/*****
/* Author: John Adams
/* Input files: C:\Documents and
Settings\e64190\Desktop\Thesis\complete
/* C:\Documents and
Settings\e64190\Desktop\Thesis\completea
/* Title: Thesis code, part 3 of 3, analysis code
/*****

options formdlm = 'j'
      nofmterr
      fmtsearch=(thesis)
      mstored
      sasstore = macros;

libname thesis 'C:\Documents and Settings\e64190\Desktop\Thesis';
libname macros 'C:\Documents and
Settings\e64190\Desktop\Thesis\macros';

/* determines the number of people in original data set that answered
sxq */
proc freq data = thesis.complete;
      tables sxq;
run;

proc freq data = thesis.complete;
      tables sxq*allog1;
run;

/* assesses normality of age */
proc univariate data = thesis.complete plots;
      var ridageyr;
      weight mec4yr;
run;

/* sorted in order to use the by statement in proc surveyfreq */
proc sort data = thesis.completea out = thesis.completea;
      by sxq;
run;

/* used thesis.completea because it contains all responses in sxq
module */
/* this gives the prevalence by sexual identity
*/
proc surveyfreq data = thesis.completea;
      by sxq;
      strata sdmvstra;
      cluster sdmvpsu;
      weight mec4yr;
      tables pir      race agec dmededuc riagendr alcohol smoke duq100
hsd010
      albumin crp bmi totchol hdl      ghemo hr      sbp      dpb;
run;

```

```

/* used thesis.completeal because it contains all resposes in sxq
module */
/* gives chisq analysis of homogeneity between categories and sex id
*/
proc surveyfreq data = thesis.completea;
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    tables sxq * (pir race agec dmddeduc riagendr alcohol smoke duq100
hsd010
                                albumin crp bmi totchol hdl ghemo hr sbp dpb) /
chisq;
run;

/* this is used for predicting probability of low albumin based on
relevant factors */
/* smoking, total cholesterol SEP (via PIR) and BMI as well as sexual
identity. */
proc surveylogistic data = thesis.completeal;
    class albumin (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model albumin = sxq smoke totchol pir bmi;
run;

/* this is used for predicting probability of low hdl chol based on
relevant factors */
/* BMI, SEP (via PIR), ethnicity (via RACE), smoking status, alcohol
use, gender and */
/* sexual id
*/
proc surveylogistic data = thesis.completeal;
    class hdl (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model hdl = sxq bmi pir alcohol riagendr race smoke;
run;

/* this is predicting probability of hight tot chol based on relevant
factors */
/* BMI, SEP (via PIR), ethnicity (via RACE), smoking status, alcohol
use, gender and */
/* sexual id
*/
proc surveylogistic data = thesis.completeal;
    class totchol (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model totchol = sxq bmi pir alcohol riagendr race smoke;
run;

```

```

/* this is used for predicting probability of high glyco-hemoglobin
based on */
/* relevant factors BMI, ace, race and total cholesterol as well as
sexual */
/* identity.
*/

```

```

proc surveylogistic data = thesis.completea;
    class ghemo (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model ghemo = sxq bmi agec race totchol;
run;

```

```

/* this is used for predicting probability of obesity based on relevant
factors */
/* race age as well as sexual identity.
*/

```

```

proc surveylogistic data = thesis.completeal;
    class bmi (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model bmi = sxq race agec;
run;

```

```

/* this is used for predicting probability of high albumin based on
relevant factors */
/* bmi, smoking status, diabetes (ghemo), high systolic and diastolic
bp and sex id */

```

```

proc surveylogistic data = thesis.completeal;
    class crp (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model Crp = sxq bmi smoke ghemo sbp dpb;
run;

```

```

/* this is used for predicting probability of high sys bp based on
relevant factors */
/* bmi, alcohol, diabetes (ghemo), smoking status, race and sex id
*/

```

```

proc surveylogistic data = thesis.completeal;
    class sbp (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model sbp = sxq bmi alcohol ghemo smoke race;
run;

```

```

/* this is used for predicting probability of high dbp based on
relevant factors */
/* bmi, alcohol, diabetes (ghemo), smoking status, race and sex id
*/

```

```

proc surveylogistic data = thesis.completeal;
  class dpb (ref = '0');
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  model dpb = sxq bmi alcohol ghemo smoke race;
run;

/* this is used for predicting probability of high hr based on relevant
factors */
/* bmi, smoking status, diabetes (ghemo), high systolic and diastolic
bp and sex id */
proc surveylogistic data = thesis.completeal;
  class crp (ref = '0');
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  model hr = sxq agec race dpb sbp;
run;

/* this dataset excludes all missing data, to construct an accurate
allostatic load */
/* only health status confounds acts as a confounder for sxq in the
model. */
proc surveylogistic data = thesis.complete;
  class allog1 (ref = '0')
    sxq (ref = 'Heterosexual');
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  model allog1 = sxq dmddeduc riagendr agec race pir hsd010;
run;

/* this is the propensity score matching model. */
/* all covariates and demographics were included in the model. */
proc logistic data = thesis.complete;
  class sxq (ref = 'Heterosexual');
  model sxq = agec riagendr dmddeduc race pir alcohol duq100 smoke
hsd010;
  output out = thesis.ppm prob=prob;
run;

%GREEDMTC (thesis, ppm, sxq, match);

/* these are for the matched analysis */

/* sorted in order to use the by statement in proc surveyfreq */

proc freq data = thesis.match;
  tables sxq;
run;

```

```

proc surveymeans data = thesis.match;
  by sxq;
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  var ridageyr;
run;

proc surveyreg data = thesis.match;
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  model ridageyr = sxq / anova;
run;

proc sort data = thesis.match out = thesis.match;
  by sxq;
run;

data thesis.match;
  set thesis.match;
run;

/* used thesis.complete1 because it contains all responses in sxq
module */
/* this gives the prevalence by sexual identity
*/
proc surveyfreq data = thesis.match;
  by sxq;
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  tables pir      race agec dmdeduc riagendr alcohol smoke duq100
hsd010
      albumin crp  bmi  totchol hdl      ghemo  hr    sbp    dpb;
run;

/* used thesis.complete1 because it contains all responses in sxq
module */
/* gives chisq analysis of homogeneity between categories and sex id
*/
proc surveyfreq data = thesis.match;
  strata sdmvstra;
  cluster sdmvpsu;
  weight mec4yr;
  tables sxq * (pir race agec dmdeduc riagendr alcohol smoke duq100
hsd010
      albumin crp bmi totchol hdl ghemo hr sbp dpb) /
chisq;
run;

/* this is used for predicting probability of high albumin based on
relevant factors */

```

```

/* smoking, total cholesterol SEP (via PIR) and BMI as well as sexual
identity. */
proc surveylogistic data = thesis.match;
    class albumin (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model albumin = sxq smoke totchol pir bmi;
run;

/* this is used for predicting probability of low hdl chol based on
relevant factors */
/* BMI, SEP (via PIR), ethnicity (via RACE), smoking status, alcohol
use, gender and */
/* sexual id
*/
proc surveylogistic data = thesis.match;
    class hdl (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model hdl = sxq bmi pir alcohol riagendr race smoke;
run;

/* this is predicting probability of hight tot chol based on relevant
factors */
/* BMI, SEP (via PIR), ethnicity (via RACE), smoking status, alcohol
use, gender and */
/* sexual id
*/
proc surveylogistic data = thesis.match;
    class totchol (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model totchol = sxq bmi pir alcohol riagendr race smoke;
run;

/* this is used for predicting probability of high glyco-hemoglobin
based on */
/* relevant factors BMI, ace, race and total cholesterol as well as
sexual */
/* identity.
*/
proc surveylogistic data = thesis.match;
    class ghemo (ref = '0');
    strata sdmvstra;
    cluster sdmvpsu;
    weight mec4yr;
    model ghemo = sxq bmi agec race totchol;
run;

/* this is used for predicting probability of obesity based on relevant
factors */

```

```
/* race age as well as sexual identity.  
*/
```

```
proc surveylogistic data = thesis.match;  
    class bmi (ref = '0');  
    strata sdmvstra;  
    cluster sdmvpsu;  
    weight mec4yr;  
    model bmi = sxq race agec;  
run;
```

```
/* this is used for predicting probability of high albumin based on  
relevant factors */  
/* bmi, smoking status, diabetes (ghemo), high systolic and diastolic  
bp and sex id */
```

```
proc surveylogistic data = thesis.match;  
    class crp (ref = '0');  
    strata sdmvstra;  
    cluster sdmvpsu;  
    weight mec4yr;  
    model Crp = sxq bmi smoke ghemo sbp dpb;  
run;
```

```
/* this is used for predicting probability of high sys bp based on  
relevant factors */  
/* bmi, alcohol, diabetes (ghemo), smoking status, race and sex id  
*/
```

```
proc surveylogistic data = thesis.match;  
    class sbp (ref = '0');  
    strata sdmvstra;  
    cluster sdmvpsu;  
    weight mec4yr;  
    model sbp = sxq bmi alcohol ghemo smoke race;  
run;
```

```
/* this is used for predicting probability of high dbp based on  
relevant factors */  
/* bmi, alcohol, diabetes (ghemo), smoking status, race and sex id  
*/
```

```
proc surveylogistic data =thesis.match;  
    class dpb (ref = '0');  
    strata sdmvstra;  
    cluster sdmvpsu;  
    weight mec4yr;  
    model dpb = sxq bmi alcohol ghemo smoke race;  
run;
```

```
/* this is used for predicting probability of high hr based on relevant  
factors */  
/* bmi, smoking status, diabetes (ghemo), high systolic and diastolic  
bp and sex id */
```

```
proc surveylogistic data = thesis.match;  
    class crp (ref = '0');  
    strata sdmvstra;  
    cluster sdmvpsu;
```

```
weight mec4yr;
model hr = sxq agec race dpb sbp;
run;

/* this dataset excludes all missing data, to construct an accurate
allostatic load */
/* only health status confounds acts as a confounder for sxq in the
model. */
proc surveylogistic data = thesis.match;
class allog1 (ref = '0')
    sxq (ref = 'Heterosexual');
strata sdmvstra;
cluster sdmvpsu;
weight mec4yr;
model allog1 = sxq dmddeduc riagendr agec race pir hsd010;
run;
```







