

SUBSTATE REGION CONTEXTS IN PREDICTIVE MODELS FOR SEXUALLY  
TRANSMITTED INFECTIONS IN THE UNITED STATES, 2012-2019

By

Claire Louise Schertzing

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Epidemiology – Doctor of Philosophy

2024

## ABSTRACT

Sexually transmitted infections (STIs) have increased steeply in the United States (US) since the beginning of the 21<sup>st</sup> century. At the same time the US population has become less religious, with growing numbers of individuals especially in the younger age groups reporting no religious affiliation. As patterns of STIs and religious belief have changed several studies exploring the relationship between these two areas have been conducted with mixed results. This dissertation adds to the existing literature by exploring a suspected predictive relationship between religion and notifiable STI outcomes at the substate level in the US.

In the first aim, predicted estimates of substate region-specific religiosity levels were derived from four religion items assessed by the 2002-2011 National Survey on Drug Use and Health (NSDUH). I then aggregated STI case numbers from 2012-2019 at the NSDUH-defined substate region-level and used the religiosity estimate derived in aim one to examine the role of religiosity in predicting STI outcomes for the three nationally notifiable sexually transmitted infections individually and in aggregate. In aim two I used this data to examine the effect of population size and age distribution of the substate regions on the ability of religiosity to predict STI outcomes. The final aim explores the effect of the addition of supplemental covariates to each of the four models first individually then together to build a final best fit model for each outcome.

In the first aim I confirmed that the four religion items assessed by NSDUH can be used to predict a single latent dimension that I have called 'religiosity'. In the second aim I assessed the ability of religiosity alone to predict STI outcomes. I then added covariates for population size and the age distributions of the substate regions in 2002-2011 into the crude model which showed that both variables improved the fit of the religiosity predictive model for the 2012-2019

STI outcomes. Finally, in my third aim I considered several additional covariates for inclusion in the model and evaluated whether they might improve the fit of the STI prediction model restricted to terms for religiosity level, population size, and age distributions. Following this I created predictive models with plausible and unlikely predictors to arrive at a best fit model for each outcome under study. While the best fit models for each outcome varied somewhat, several covariates qualified for inclusion in all models including the substate region population, the proportion of 26- to 35-year-olds, and the extra-medical use of a set of drug and medicine subtypes identified in NSDUH modules on this topic.

Taken together, the results of this dissertation point towards three conclusions. First, religiosity may be an important predictor of STI outcomes in the US and warrants further consideration. Second, while this project does not draw causal conclusions about the relationship between religiosity and STI outcomes it suggests that a modeling approach like the one used here may be useful for future study and to better target public health programming to high-risk populations. Finally, several covariates assessed as part of this project might prove to have importance in future predictive models for STI occurrence in community and substate regions of the United States. Future studies should build upon these models by considering additional covariates as well as attempting to replicate these results with alternative or expanded data sources.

To everyone who believed in me when I couldn't believe in myself:  
You were right. This is for you

## **ACKNOWLEDGEMENTS**

I would like to thank the United States Substance Abuse and Mental Health Services Administration Center for Behavioral Health Statistics and Quality for sponsoring the National Survey on Drug Use and Health and making the data publicly available for use. In addition, I would like to thank the Centers for Disease Control and Prevention National Center for HIV, Viral Hepatitis, STD, and TB Prevention for supporting the AtlasPlus database and making data available for public use. Without CDC and SAMHSA this study would not be possible.

This project would not have been possible without the efforts of many. Special acknowledgment goes to my committee: Xiaoyu Liang, Paul Quinlan, Sue Grady, and especially Jim Anthony who guided me to and through this project. Many thanks to David Barondess for guiding me and answering my unending questions over the past four years. Thank you furthermore to all of the professors in the department who were generous with their time and knowledge during my education. A special thank you belongs to Santiago Salinas, my undergraduate mentor and the person who first got me excited about research. Finally, to my friends and family, endless thanks for your support throughout my educational journey. While I may not express it frequently your support has not gone unnoticed or unappreciated.

## TABLE OF CONTENTS

LIST OF ABBREVIATIONS.....	vii
Chapter 1. Introduction and Specific Aims.....	1
Chapter 2. History, Background, and Significance.....	3
2.1 Overview.....	3
2.2 History and Background.....	3
2.3 Sexually Transmitted Infections in the United States.....	14
2.4 Significance.....	19
Chapter 3. Materials and Methods.....	21
3.1 Overview.....	21
3.2 Aim 1.....	22
3.3 Aim 2.....	24
3.4 Aim 3.....	28
Chapter 4. Results.....	32
4.1 Aim 1.....	33
4.2 Aim 2.....	34
4.3 The Issue of Age.....	39
4.4 Effect of Individual Predictors on Religiosity.....	40
4.5 Other Plausible Predictors.....	41
4.6 Unexpected Predictors.....	42
4.7 Summary Model and Sensitivity Analyses.....	43
4.8 Post-Estimation Exploratory Data Analysis.....	43
Chapter 5. Discussion and Limitations.....	44
5.1 Aim 1.....	44
5.2 Aim 2.....	45
5.3 Aim 3.....	48
Chapter 6. Future Directions.....	52
REFERENCES.....	54
APPENDIX A: SUPPLEMENTAL TABLES AND FIGURES.....	61
APPENDIX B: PROGRAM CODE USED TO DERIVE THIS STUDY.....	74

## LIST OF ABBREVIATIONS

ACA	Affordable Care Act
CDC	Centers for Disease Control and Prevention
CFA	Confirmatory Factor Analysis
CSTE	Council of State and Territorial Epidemiologists
DALY	Disability Adjusted Life Year
DAS	Restricted Use Data System
GLM	Generalized Linear Model
HPV	Human Papilloma Virus
IQR	Interquartile Range
LR	Likelihood Ratio
MFP	Multivariable Fractional Polynomial
NBREG	Negative Binomial Regression
NNDSS	Nationally Notifiable Disease Surveillance System
NSDUH	National Survey on Drug Use and Health
PID	Pelvic Inflammatory Disease
RVF	Residual versus Factor
SAMHSA	Substance Abuse and Mental Health Services Administration
SEM	Structural Equation Model
SIECUS	The Sexuality Information and Education Council of the United States
STI	Sexually Transmitted Infection
US	United States
USPSTF	United States Preventive Services Task Force

## Chapter 1. Introduction and Specific Aims

Health departments across the United States (US) from the local to national level have been dealing with rising numbers of sexually transmitted infections since the early 2000s. Disparities in funding and other resources as well as prevailing political and organizational priorities limit the ability of many organizations to adequately address these diseases in the populations they serve. How and why public health resources are allocated is an issue policy makers and public health professionals must make frequently. It is impossible to address every health condition affecting the US population at once, and there is no way to allocate the available funds and resources that will make everyone happy. This is especially true in the case of stigmatized conditions like sexually transmitted infections which can elicit strong feelings and opinions from policy makers and the public alike.

While it can be a polarizing topic there is no denying that sexually transmitted infections can have serious health implications and that cases are increasing across the country. Numerous factors across multiple levels of society affect STI risk including: healthcare seeking behavior, sexual activity and sexual networks, social values, health system policies, and sexual and racial identity among many others (Hamilton et al., 2023).

Not all these factors that have been shown to have an association with STIs have a clear positive or negative effect on STI risk. Religion is one of these factors. Studies published on the association between religion and STIs have shown both positive and negative effects of religion on STI risk. This dissertation research seeks to determine whether substate region-level religiosity can be used to predict STI prevalence at the substate level in the US, and to provide a stepping-stone for future research to clarify the relationship between religion and STI risk in the US.



An important departure of this dissertation research from currently published work on religion and STIs is the linkage of two publicly available data sets as well as the modeling focus of the work. Publicly available data is an important resource for research, through this dissertation I have created a crosswalk between two large data sets that expands the utility of the data. Additionally, prior studies published on the religion – STI relationship have largely assessed association. In contrast, I have developed a series of predictive models which can be applied forward to predict STI prevalence at the substate level. While the importance of religiosity to predicting STI outcomes is important for public health, it is only one piece of the puzzle. This project also seeks to discover additional predictors of STI outcomes beyond religiosity. These predictors together may be helpful for health departments to plan resource allocation and target public health programs to high-risk groups to achieve greater efficacy.

This dissertation research project has three specific aims which are as follows:

1. To investigate the dimensions of religiosity at the substate region level of the United States as manifest in a set of four standardized survey items on this topic.
2. To measure the substate region-level's position on the just-described religiosity dimension for 2002-2011 and to estimate the degree to which that position might help account for the occurrence of notifiable sexually transmitted infections (STI) in the US.
3. To estimate the degree to which the fit of the religiosity model for occurrence of notifiable STI developed under Aim #2 can be improved via addition of other theoretically specified covariate predictors from the 2002-2011 interval, as derived from the US NSDUH substate region dataset for 2002-2011.

## Chapter 2. History, Background, and Significance

### 2.1 Overview

Sexually transmitted infections have plagued humanity since antiquity. Ancient religious and lay texts include descriptions of diseases of the genitals believed by scholars to be STIs. As with most ailments early societies attributed STIs to supernatural causes such as divine retribution or the work of dark magic. Priest healers or shamans served the physical and spiritual health of the population, suggesting a mixture of physical cures as well as prayers or rituals to alleviate suffering. The medieval period and emergence of Christianity changed the way health and medicine were viewed. Disease was punishment for sin, cures focused on the healing of the soul as opposed to physical intercession and diseases associated with sin such as STIs became more heavily stigmatized.

The Renaissance brought about the beginning of secular medicine with medical education being offered outside the church for the first time. This division of religion and medicine separated medical practitioners from religious institutions, but it did not separate religious beliefs and practices from health. Even today we see examples of people using prayer and other religious practices to combat sickness. Furthermore, certain religious proscriptions relate directly or indirectly to health, such as the practice of abstinence outside of marriage or the avoidance of touching non-familial members of the opposite sex (an orthodox Jewish concept called *shomer negiah*).

### 2.2 History and Background

#### 2.2.1 A Brief History of Sexually Transmitted Infections in Antiquity

When sexually transmitted infections (STIs) emerged is a topic of debate, however, evidence of diseases whose presentations are suggestive of STIs existence from as early as ancient Mesopotamia (Gruber et al., 2015). Cuneiform tablets describing vaginal and urethral

discharge as well as pustules on the genitals have been discovered during excavation of their ancient cities. Mentions of 'heat' associated with difficulty passing urine as well as 'secretions' that cause suffering can be found in the Eber's papyrus, leading some to suggest that these symptoms may have been caused by gonorrhea or chlamydia. The Bible is occasionally cited as including early accounts of likely STIs. Leviticus 15 for example describes rules for cleanliness as well as physical and spiritual purification for men and women when they have an 'unusual bodily discharge', which is believed by some scholars to be suggestive of gonorrhea or chlamydia infection (Flemming, 2019). The painful sores with which Satan afflicts Job from head to foot in Job 2:7 has also been suggested to be a description of syphilis. These are just two of several descriptions of potential STIs which exist in the Bible, however, the descriptions and accuracy are questionable as the Bible was written by lay-people rather than physicians (Gruber et al., 2015; Flemming, 2019).

Medicine moved away from religious superstition thanks to Hippocrates, his written corpus developed the medical vocabulary and includes significant advancements in medical and scientific thinking. According to Hippocrates four humors controlled health: blood, yellow bile, black bile, and phlegm. An imbalance in any of the humors was the cause of disease in individuals and must be treated to restore health. Frequent intercourse was recommended to reduce the level of phlegm in the body, a fever was caused by an excess of blood in the body and so required bloodletting to be cured, for example. In his writings Hippocrates describes several symptoms that could be attributable to acute gonorrhea as well as ulcerative lesions and growths on male and female genitalia believed to be herpes and genital warts (Gruber, 2015).

Despite numerous writings from Greece and Rome regarding diseases associated with sexual contact the descriptions do not entirely align with the modern presentation of sexually

transmitted infections. Symptoms including vaginal and penile discharge, sores, painful urination, weakness, fever, pain in the genitals, and emaciation are all present in medical texts from the period; however the symptoms are not grouped together as would be suggestive of gonorrhea or syphilis. Instead, symptoms are described individually in relation to other diseases such as consumption, strangury, and *tuphos*. A disease termed *Gonorrhoeia* does exist in Roman texts, including medical texts summarizing conditions previously known by the Greeks and the writings of the philosopher Celsus, but the described infection which includes discharge from the genitals and general wasting does not meet the modern case definition of gonorrhea (Flemming, 2019). Due to these descriptions some scholars have suggested that while sexual diseases existed in the ancient world, our modern pathogens developed sometime later.

### 2.2.2 A Brief History of Religion

Throughout history there have been a vast number of religions and religious expressions across cultures, time periods, geographies, and groups. As both a psychological and social construct it follows that some universalities exist within and between religions but also that some aspects vary by culture, location, time, and denomination. Broadly, religion, derived from the Latin *religare* meaning ‘to bind’, brings people together in community as well as bringing them closer to their God through a set of common practices. Emile Durkheim states in *The Elementary Forms of the Religious Life* that the earliest human representations of the world and our place in it was based in a ‘speculation of the divine’ which he considered primitive religion (Longhofer & Winchester, 2016; Durkheim, 1915). Durkheim argues that religion began nowhere, that instead it has existed since humans first developed a concept of themselves as part of the world. This early religion, a “speculation upon divine things”, was borne from searching for an explanation

for phenomena beyond human understanding, that which was magical or supernatural (Durkheim, 1915).

Whether religion has existed since humans became cognizant of their place in the world as Durkheim suggests or whether it developed later, it will likely never be possible to pinpoint the origin of religion. However, it is entirely possible to trace a trajectory of religious belief from its primitive to modern forms. As societies grew and larger groups of people began to live together, religion became more organized, a nebulous 'divine' gave way to greater organization with deities, ceremonies, and the need for specific religious personnel. Early organized religions like those of the Egyptians, Greeks, and Aztecs were among the first to name their gods and emphasized obedience to the deities and offerings to curry favor (Baumard & Boyer, 2013). The gods controlled all aspects of life, their existence offered a consistent mythical explanation for natural phenomena and a target to whom prayers and offerings could be made to improve one's life (Risse, 1993).

More recently, several religions with specific moral prescriptions have emerged in which the concept of the deity or higher power is linked to specific ethical or moral ideals (Baumard & Boyer, 2013). For Christians and Jews for example, rather than gaining favor through offerings an individual's success or failure in following certain moral doctrines determines whether one receives protection from their God. The long history of religion has led to a great number of different belief systems, traditions, and definitions of religion both within the literature and between individuals. Due to its simultaneously collective and individual nature, a wide range of expressions of belief are encompassed within 'religion'. While certain group values or practices may be easy to spot, at the individual level the frequency of participation or (non)adherence to traditions and norms varies. Thus individual 'religiosity' may be overtly visible (e.g., attending

church) but it also may be private (e.g., a feeling of connection to God). Furthermore, while one person may define their religion by identifying the religious denomination to which they belong, another may point to the rituals they practice and a third may describe an ethical or moral belief system to which they ascribe.

Definitions of religion differ not only based on personal and group identity but also by the intended use of the definition. Numerous papers have been published advocating for the need to clearly define ‘religion’ in studies seeking to examine its effect on any number of human conditions and experiences (Miller & Thoresen, 2003; Saroglou, 2012; Moon, 2023). None, however, has been able to propose a definition that would be broadly applicable.

Three of the four questions pertaining to religion asked by the NSDUH survey between 2002 and 2011 are largely focused on the outward expression of religious belief. Attending religious services and choosing to associate with individuals of the same religion are outwardly visible signs of an individual’s beliefs. Basing one’s decision making on their religion may lead to outward expression of some of the moral tenets of religious identity (e.g., voting for a political candidate who shares your beliefs). The last item, whether one’s religious beliefs are an important part of their life, is a less outwardly visible measure but is likely deeply tied to an individual’s response to the other three questions. Because these four questions tap both outward and potentially private facets of an individual’s religiosity, the Handbook of Religion and Health (2<sup>nd</sup> ed.) definition most fully captures ‘religiosity’ as reflected in the NSDUH survey. This definition is both comprehensive and for the purpose of this study best reflects the vast spectrum of human religious behavior.

*“[Religion] Involves beliefs, practices, and rituals related to the transcendent, where the transcendent is God, Allah, HaShem, or a Higher Power in Western religious traditions, or*

*to Brahman, manifestations of Brahman, Buddha, Dao, or ultimate truth/reality in Eastern traditions. This often involves the mystical or supernatural. Religions usually have specific beliefs about life after death and rules about conduct within a social group. Religion is a multidimensional construct that includes beliefs, behaviors, rituals, and ceremonies that may be held or practiced in private or public settings, but are in some way derived from established traditions that developed over time within a community. Religion is also an organized system of beliefs, practices, and symbols designed (a) to facilitate closeness to the transcendent, and (b) to foster an understanding of one's relationship and responsibility to others in living together in a community.” (Koenig, King & Carson (Eds.), 2012, in Koenig, 2012).*

While we struggle to perfect a modern definition of religion, religion at its inception was far removed from its contemporary manifestations. Numerous theories of religious development have been proposed, three of which I will describe here. The first, to which scholars including Emile Durkheim, Mircea Eliade and Ioan Couliano align themselves, states that religion began with the earliest bipedal hominins, ascribing to them an intelligence and imagination like modern humans (Wunn, 2000). Durkheim illustrates this idea in *The Elementary Forms of Religious Life* using the modern Aboriginal tribes of Australia as an example of what he believes the ‘uncivilized’ beliefs and religion of early humans would have been like (Dow, 2006; Durkheim, 1915). However, while aspects of Aboriginal life may be considered ‘primitive’ in comparison with Western society they have nonetheless been affected by their history and cultural evolution in the same way as any other society. Comparison of primitive and modern hunter-gatherer religions is impossible, let alone modern hunter-gatherer religions and mainstream religions, as religious ideas change over time the same way belief systems do (Grafton, 1945; Wunn, 2000).

The second idea proposes a still prehistoric but significantly later development of religion than the first. This theory points to the development and use of symbols as the point at which religious thought came into existence. Symbols have been found in human settlements from as far as 100,000 years ago, however, it is not until between 65,000 and 35,000 years ago that many anthropologists would say some form of religious thought likely existed. The explosion of symbolic evidence during this period, including cave paintings and carved figurines combining human and animal features provide the strongest evidence for the existence of some early religion (Wunn, 2000; Culotta, 2009).

The third school of thought takes a more Darwinian approach, noting that early bipedal hominins did not have the brain capacity for abstract or symbolic thought, a commonly accepted prerequisite for religious thinking (Henning, 1898; Wunn, 2000; Dow, 2006; Culotta, 2009). Development of the cognitive architecture required for higher level thought was necessary for early humans to create the symbols and develop burial practices that have been pointed to as evidence for early religion (Watts, 2020). While evolution of the cognitive architecture occurred for survival scholars have shown that religious thought piggybacks on those same processes, for example, a predisposition to ascribe human characteristics to inanimate objects or phenomena. Assuming a 'who' (e.g. man, animal) is responsible for a sudden noise, for example, puts one on alert in a way assuming a 'what' (e.g. wind) does not. What was likely developed as a survival mechanism and is now a hold-over from natural selection supports religious thought, allowing humans to conceptualize a higher power (Watts, 2020). The provision of agency to these objects or circumstances when extended to an omnipotent higher power then naturally gives them the same thought process as humans, both good and bad. One or more omnipotent deities may also provide a way to control groups of individuals, common belief may increase cooperation and



improve group survival while also providing a measure of protection from bad behavior due to ‘surveillance’ from a higher power (Culotta, 2009; Dow, 2006).

Outward expressions of religion such as those measured by the four religion variables collected by NSDUH serve as a signal of group identity. Shared beliefs bring individuals together while visible manifestations of those beliefs prove to other members of the community that an individual is an upstanding member of the group. Corrupt or atheistic actions may then not only be watched and punished by an omnipotent deity but also the group itself, creating a system in which individuals are both morally obligated by their religion to act appropriately and socially obligated to maintain their in-group status.

### 2.2.3 Intersections of Religion and Health

Religion and spirituality have influenced society, history, and daily life for thousands of years. Belief systems shape morality and ethics, culture, family and interpersonal dynamics, and shapes an individual’s ideas about the world and oneself (Camino-Gaztambide, Fortuna, Stuber, 2022). For much of history religion was medicine and vice versa. Early humans seeking to understand their existence and alleviate suffering looked to healers who served as agents between the divine and the clan, working to discern what invisible forces were affecting the group and helping to restore harmony between the human and divine (Levin, 2020). These early shaman healers like the *wab* (priest healer) of Egypt and *wu* (diviner and wizard) of China were consulted when domestic healing failed, or a problem was deemed too serious to be dealt with at home (Risse, 1993). As greater anatomical and clinical understanding was developed and shared, the role of shaman healers was slowly overtaken by more ‘professional’ healers with a wider understanding of health and disease. These professional healers became more specialized as civilizations expanded. Increased conflict between populations for example necessitated the

development of nutrition and fitness specialists as well as surgical specialists to prepare and heal individuals engaged in battle (Risse, 1993).

Beyond mechanical causes such as battle wounds, physical ailment was attributed to a variety of causes in the ancient world, including divine retribution, the work of the devil, dark magic, and natural causes (Levin, 2020). Treatments for disease were often a mixture of religious ritual and potions or salves, frequently administered by clergy-physicians. The mixture of etiological ideas about disease led to numerous types of healers. Priests, magicians, folk-healers, and physicians all served the health of the population in amicable competition, seeking the services of one type of healer did not exclude the use of another (Levin, 2020). Some of the most well-known advances in medical knowledge come from Hippocrates and his students and followers. The Hippocratic corpus, including the theory of bodily humors originally proposed by Hippocrates and later developed by Galen, heavily influenced medical treatment by physicians and healers for centuries. Balancing the four humors (yellow bile, black bile, phlegm, and blood) was considered the key to health. An overabundance of any of the humors required purgative treatment such as bleeding or vomiting. Despite centuries of use, the advances made by Hippocrates and Galen and the knowledge developed based on their ideas during the Greek and Roman empires was lost during the Dark Ages when Christianity became the dominant religion and shifted the emphasis from physical health to spiritual salvation (Risse, 1993). Sickness was considered punishment for sin, the church emphasized cleansing of the soul through prayer, exorcism, and miracles as a cure for disease rather than the use of medical knowledge. Treatment of the sick became a focus for religious organizations with several hospitals being created and staffed by monastic orders or clergy (Risse, 1993). Whether the disease was believed to be due to the wrath of a deity or promiscuity and loose morals, individuals infected with these conditions

have been stigmatized and shamed since antiquity. As Christianity took over in many places during the Middle Ages, chastity was increasingly seen as good and ‘godly’ while sex outside of marriage was sinful, dirty, and unhealthy. Prostitution was commonly blamed for outbreaks of STIs. Faced with an inability to cease prostitution to prevent disease from spreading, laws were enacted to regulate the practice as early as 1161 in England and 1256 in France.

Through the medieval period, what little medical training existed was concentrated in religious universities which drew their authority to license physicians from the Pope. Forbade by the church from dealing with procedures involving bloodletting or the possibility of fatalities, physicians were academics while surgery was left to other types of healers such as barber-surgeons (Gelfand, 1993).

The Renaissance brought about the rediscovery and challenge of Galenic anatomic knowledge and a growing tie between physicians and surgeons. Renaissance humanism, based in the study of classical antiquity, placed an emphasis on reasoning and experimentation that led to advances in philosophy, art, the sciences, and religious thought which lasted centuries (Kristeller, 1978). During this period healer guilds gained power and political leaders began to patronize physicians. Just as the Medici family patronized artists such as Michelangelo and Botticelli, so too did kings patronize physicians and medical personnel. In France Ambroise Pare, a barber-surgeon today considered the father of surgery and forensic pathology, was the royal surgeon for not one but four French rulers (Hernigou, 2013). Medical patronization by political leaders like that of Pare by the French monarchy represented a move away from church-controlled licensure of physicians and towards state control (Gelfand, 1993; Koenig, 2012; Tansey, 1993). While physical health care became more secular during this period, mental health care remained largely

within the purview of religious institutions until the end of the French Revolution some 250 years later (Koenig, 2000; Koenig, 2012).

The separation of medicine and medical treatment from religious rites and institutions has largely held over the last several centuries. However, religion re-entered medicine as a potential cause or contributing factor of disease, as opposed to a cure, in the 19<sup>th</sup> century. Published in 1897 by French sociologist Emile Durkheim, *Suicide: A Study in Sociology* explored social and natural phenomena related to suicide. Psychology, anthropology, religion, family, social crises, history, and education are all among the phenomena explored in relation to suicide in his work (Durkheim, 1897). Religion, according to Durkheim, serves as a socially cohesive force which prevents suicide. He compares Catholicism and Protestantism to show how the Catholic faith with its unified traditions and hierarchical organization prevents the religious interpretation permitted in Protestantism, thereby promoting obedience, preventing schisms, and reducing suicide (Durkheim, 1897). The increasing importance of individual thinking, what Durkheim terms ‘egoism’, leads to the breakdown of the strongly integrated social groups he deems necessary to prevent suicide. Thus, religion, or the lack thereof, contributes to ‘egoistic suicide’. Following Durkheim’s recognition of the link between religion and health numerous studies have been published examining the association between religion and various health conditions including mental health diagnoses such as depression, schizophrenia, and anxiety, as well as cancer, heart disease, dementia, kidney disease, COVID-19, and the focus of this dissertation, sexually transmitted diseases (Adams et al., 2020; Agli et al., 2015; Awaworyi Churchill et al., 2021; Hemmati et al., 2019; Koenig, 2000; Moons et al., 2019; Nair et al., 2020; Sisti et al., 2023; Weber & Pargament, 2014).

The association between religion and health is well documented, however the association is not unidirectional as religion and spirituality have been shown to have both positive and negative effects on health. Positive effects of religion include better mental health such as less depression, lower stress and anxiety, and greater social support, slower cognitive decline in individuals with Alzheimer's disease, reduced likelihood of smoking, and lower risk of coronary heart disease (Agli et al., 2015; Koenig, 2012; Weber & Pargament, 2014). However, religion can also act as a barrier to health when religious or spiritual beliefs conflict with medical advice or when individuals hold punitive views of God and fear punishment or feel guilt for not living up to a standard (Weber & Pargament, 2014). Religious beliefs have been linked to withholding medical care to an extent which constitutes child abuse in several cases, as well as refusal of potentially lifesaving care such as blood transfusions, prenatal care, and vaccination (Koenig, 2000).

In the case of sexual health, religion has shown to be protective in delaying initiation of sexual activity, being more likely to decline sexual advances, and reducing the number of sexual partners, all of which reduce the risk of sexually transmitted infections. However, studies have also shown that highly religious individuals are less likely to seek sexual and reproductive health services than their less religious counterparts (Hall et al., 2012). Additionally, it is important to note that unmarried religious individuals may be less likely to receive information on safe sexual practices and therefore at higher risk of disease (Awaworyi Churchill et al., 2021).

### 2.3 Sexually Transmitted Infections in the US

Case numbers for all three infections have risen steadily over the last several decades, leading to increasing costs to the healthcare system as well as disability and death in infected individuals. Various efforts to reduce the number of infections have been suggested and

implemented, but with so many factors contributing to STI risk they have yet to be successful. One method that has been suggested to reduce the burden of disease is to provide comprehensive sexual health education in schools. As of 2022 however, sexual health care and sexual health education vary widely across the United States.

The Sexuality Information and Education Council of the United States (SIECUS) keeps up-to-date information on sexual health policy in the US and its territories. According to their 2022 report only 29 states and the District of Columbia require sexual health education be taught to students as part of the curriculum. When sexual health education is provided, 30 states require that instruction stresses abstinence, and 13 states do not require that the information provided be age appropriate, evidence based, or medically accurate (SIECUS, 2022). In states where sexual health education (“sex ed”) is not mandatory school districts have control over whether education is provided and what type of information is shared. Additionally, all but two states have either opt-in or opt-out policies for sexual health education, meaning parents can decide to remove their child from participating in sex ed (SIECUS, 2022). This leads to a great deal of heterogeneity in the information provided to students across school districts, states, and the US.

Sexual health care is also highly variable within the US. Health care access was expanded with the passage of the Affordable Care Act (ACA), including expanded access to several sexual health-specific services for individuals. New individual and small group insurance plans as well as Medicaid expansion plans now have co-pay free access to STI services included as part of the Grade A and B preventive services recommended by the US Preventive Services Task Force (USPSTF) (Leichliter et al., 2016). Additionally, individuals in those states which did not expand Medicare eligibility are also able to access these services if they have a new private insurance plan through the Marketplace. While this expansion is helpful in expanding STI services many

individuals remain uninsured as they fall in the ‘coverage gap’, meaning they are ineligible for Medicare but do not qualify for Marketplace subsidies due to incomes below the federal poverty level (Rudowitz, Drake, Tolbert, and Damico, 2023).

### 2.3.1 Sexually Transmitted Infections and Identity

Expanded insurance coverage is not sufficient to help curb rising STI rates in the US. Increased access to testing and treatment does not address the social stigma which often surrounds these infections, nor does it address the individual and interpersonal factors affecting STI risk. These include sociodemographic factors, sexual behavior, mental health, substance misuse, and the influence of parents, peers, providers, sexual partners, and many others (NASEM, 2021).

Individuals who identify as part of a racial/ethnic or sexual minority experience a disproportionate burden of STIs in the United States. Poverty, social and sexual networks, racism and discrimination, as well as a history of unethical medical experimentation on these groups are only some of the factors likely driving the observed disparities. Rates of infection are 5 to 7.7 times as high in Black men compared to white men across the three major STIs with slightly smaller but still pronounced differences for Hispanics/Latinos and native populations (Lieberman et al., 2020). Women who identify as lesbian or bisexual are at greater risk of becoming infected with an STI than their heterosexual counterparts, with the greatest risk being in Black (OR 6.43) and Hispanic (OR 2.05) bisexual women (White bisexual women OR 1.77) (Mojola & Everett, 2012). White gay men (OR 8.3) and Black men of all orientations (heterosexual OR: 2.91, bisexual OR: 5.97, gay OR: 3.2) are more likely to have an STI than their White heterosexual counterparts (Mojola & Everett, 2012). As indicated by the data, individuals with multiple minority identities bear the greatest risk of STI in the US. Large increases in STI cases from

historic 2000 lows coupled with the heavy burden of disease born by racial and sexual minority groups points to a need for more focused public health measures to halt the current epidemic.

### 2.3.2 Symptoms and Background

An estimated 26 million STIs are acquired each year in the United States, almost half of which occur in individuals 15-24 years of age (CDC, 2021). The sexually transmitted diseases that are the focus of this dissertation are all bacterial infections. All three are curable diseases, however, there is growing concern over the rise of antibiotic resistant strains of the bacteria, especially gonorrhea.

#### 2.3.2.1 Chlamydia

*Chlamydia trachomatis*, the causative agent of chlamydia infection, is the most recent of the notifiable STIs to have been identified. Identified in 1972, it is the most frequently diagnosed STI with an estimated global prevalence of just under 56.4 million cases in 2019, leading to 160,000 DALYs and 972 deaths (IHME, 2015). In the US chlamydia caused 1.6 million new infections in 2021 alone (CDC, 2021). While infection is frequently asymptomatic men and women who contract chlamydia may display symptoms including abnormal vaginal or penile discharge, a burning sensation when urinating, or pain, discharge, or bleeding from the rectum if the infection is in the anus. *C. trachomatis* infection initially affects the cervix and can spread into the upper reproductive tract, causing pelvic inflammatory disease (PID). PID has the potential to cause severe damage to the reproductive tract including the uterus and fallopian tubes, leading to infertility and pelvic pain in affected women. Pregnant mothers who are infected may experience pre-term delivery and can pass the infection during childbirth leading to ophthalmia or pneumonia in the infant (CDC, 2023a).



### 2.3.2.2 Gonorrhea

Gonorrhea is the longest identified of the notifiable STI. The causative bacterium, *Neisseria gonorrhoeae*, bears the name of the man who discovered it in 1879, Albert Ludwig Sigismund Neisser. Believed to be the same disease as syphilis until 1838, it is fitting that data collection on both gonorrhea and syphilis began in the same year. Both diseases were added to the list of nationally notifiable diseases in 1944, just in time to see steep declines in disease due to the discovery of penicillin. *Neisseria gonorrhoeae* infection is typically asymptomatic. Men and women who are infected may experience abnormal genital discharge or anal itching and soreness depending on the site of the infection. Like chlamydia, when left untreated gonorrhea may lead to PID and epididymitis in women and men respectively, both of which may lead to infertility. In extreme cases gonorrhea may spread to the bloodstream to cause a disseminated gonococcal infection which can be fatal. Gonorrhea may be transmitted vertically, contact with an infected mother's genital secretions during childbirth may lead to conjunctivitis, infection of the joints, or life-threatening blood infection in infants (CDC, 2023b). In 2019 the global prevalence of gonorrhea was estimated at 20 million cases, causing 2,963 deaths and 163,000 disability-adjusted life years (DALYs) (IHME, 2015). In the United States over 700,000 cases were reported to CDC in 2021 (CDC, 2021).

### 2.3.2.3 Syphilis

Despite the long and fearsome history of syphilis, the causative bacteria, *Treponema pallidum*, was not identified until 1905. Syphilis accounts for a relatively small number of global cases at just over 31 million in 2019 but has the greatest death toll of any of the infections discussed here, 83,682 deaths and 7.4 million DALYs (IHME, 2015). In the US 176,000 cases were reported in 2021 (CDC, 2021). When left untreated infection follows a four-stage

trajectory, moving through primary, secondary, latent, and tertiary infection. Primary infection is characterized by the emergence of one or more painless ulcers on the genitals approximately three weeks after infection. Secondary infection begins six to eight weeks after ulcers resolve with the appearance of additional symptoms which may include headache, fever, or rash especially on the torso or hands and feet. Latency is entered when these secondary symptoms disappear and lasts until the individual dies or in rare cases moves to tertiary infection when the disease can cause life threatening neurological and cardiac complications (CDC, 2023c). The disease is generally considered to be ‘active’ or infectious through the early latency period which may last up to two years after initial infection. Congenital transmission risk is greatest when the mother is in the early stages of disease. Primary maternal infection poses a 50% risk of transmission while early latent infection has been shown to have a transmission risk as high as 83% (Stafford et al., 2019). Transmission may occur as early as the first trimester, leading to fetal or infant death, or birth with complications including low birthweight, pre-term birth, jaundice, and physical and mental developmental disabilities (Peeling et al. 2017; CDC, 2023c).

#### 2.4 Significance

Sexually transmitted infections (STIs) are a major public health concern globally. In 2019 alone there were an estimated 1.29 billion STIs which contributed to almost 90,000 deaths and 8.57 million DALYs (IHME, 2015). While many STIs are curable, untreated disease poses a threat to reproductive and neonatal health and large numbers of infected individuals represents a high cost to health systems (Zheng et al., 2022).

In the United States STI incidence has been increasing over the last two decades. Since 2017 alone gonorrhea has increased 28% and syphilis by 74% (CDC, 2021). The heaviest burden of which falls on youth and young adults who account for more than 50% of cases of many of

these diseases. Perhaps most concerning is the 203% increase in congenital syphilis that has been reported as it led to more than 220 infant deaths and stillbirths in 2021 alone (CDC, 2021; NASEM, 2021). Furthermore, the high prevalence of asymptomatic infections indicates that the problem is likely much worse than we can see. Combined with increasing antibiotic resistance of some of the bacteria that cause these infections it is imperative that we place a greater focus on controlling and reducing STIs.

As proposed, my dissertation is designed to assess religiosity as a latent trait manifest in four religion survey questions used by the NSDUH survey and to explore whether this latent measure of substate region-level religiosity can be used to predict STI outcomes. Additionally, it will explore additional covariates which may be used to improve the predictive ability of my models and may provide a more detailed look at the drivers of STI outcomes at the substate level. If the aims of this project are achieved, it will serve as an important step forward in building models to predict STI outcomes which can eventually help public health practitioners better target resources to the highest risk communities.

## Chapter 3. Materials and Methods

### 3.1 Overview

To understand how religiosity might impact the occurrence of nationally notifiable STIs in the United States (US) we need a valid measure of religious belief as well as substate level parameters which may improve the fit of our predictive models. By aim, this chapter will outline the methods I used to:

1. Investigate the existence of a latent religiosity dimension at the substate region level of the US as manifest in a set of four NSDUH survey items on the topic.
2. Measure the substate-level position on the religiosity dimension and build predictive models to assess the ability of religiosity to predict the occurrence of STI cases at the substate level in the US
3. Assess the fit of crude models developed under Aim #2 and estimate model fit improvements after the addition of additional covariate predictors derived from the NSDUH 2002-2011 dataset.

#### 3.1.1 Details on IRB Approval, Recruitment, and Participation Levels

The current study was determined by the MSU IRB as not human research on 10/16/2023. Proof: STUDY00009821.

Overall interview participation levels in the NSDUH are between 65%-72%. See table 1 for the sample size, response rates, and overall participation rates in the NSDUH for each year under study.

**Table 1.** NSDUH sample size and survey participation for all years included in this study

Year	Final Sample	Wtd. Screening Response Rate %	Wtd. Interview Response Rate %	Overall Survey Participation %
2002	68,126	91	79	72
2003	67,784	91	77	70
2004	67,760	91	77	70
2005	68,308	91	76	69
2006	67,491	90.23	74.21	67
2007	67,377	89.07	73.87	66
2008	67,928	88.62	74.24	66
2009	68,007	88.4	75.56	67
2010	67,804	88.42	74.57	66
2011	70,109	86.98	74.38	65

### 3.2 Aim 1

#### 3.2.1 Study population

This study is an analysis of publicly available data from the NSDUH 10-year substate restricted use data set collected between 2002 and 2011 and provided by the Substance Abuse and Mental Health Services Administration (SAMHSA). For this study I used responses from four religion survey items collected from participants in 383 NSDUH-defined substate regions in the United States.

#### 3.2.2 National Survey on Drug Use and Health Restricted-Use Data

The NSDUH survey is a cross-sectional survey which uses multi-stage probability sampling to obtain a sample population that is representative of the civilian, non-institutionalized US population ages 12 and over. Between 2002 and 2011 the NSDUH survey was administered in-person by trained interviewers using computer assisted interviewing methodologies to protect

the privacy and confidentiality of respondents. A more detailed description of the relevant methods and procedures is available elsewhere (SAMHSA, 2010).

Data from the survey available in the restricted use data system (DAS) does not show nor allow for the generation of unweighted frequencies of any variables to further protect the confidentiality of the survey participants. The data has been subsampled from the full survey population and has a revised calculated analysis weight to ensure the population is representative over the 10-year period. For three measures of religion (importance: “*my religious beliefs are very important*”, friendship: “*it is important that my friends share my religious beliefs*”, and decision making: “*my religious beliefs affect my decision making*”) participants were asked to respond based on a 4-point Likert scale ranging from strongly agree to strongly disagree. For attendance respondents were asked to recall how many times they had attended religious service in the past 12 months, excluding weddings, funerals, and other special occasions.

### 3.2.3 Confirmatory Factor Analysis

To test whether the four religion variables collected by NSDUH were in fact good predictors of a latent religiosity trait, a confirmatory factor analysis (CFA) was conducted using the structural equation modeling (SEM) command in Stata (release 18; StataCorp, 2023). Substate-level values for each of the four religion items downloaded as .csv files from the DAS system were brought first into excel and cleaned before being imported into Stata and merged to create the dataset for analysis. Due to the nature of the restricted-use data used the 10-year analysis weight was automatically applied to the results generated within the online DAS system, eliminating the need to apply a survey weight to the data during later analysis.

The substate region-level values of the latent trait religiosity measured in Aim #1 cannot be used in later modeling as Stata does not support the use of latent variables outside of the SEM

commands. To address this I created a new variable ('relighat') which contains a factor score with predicted values of the latent variable which will be used going forward.

### 3.3 Aim 2

#### 3.3.1 Study Population and Sample

For this study the population included the same representative non-institutionalized civilian population assessed by NSDUH between 2002 and 2011 as the previous aim. However, it also includes US residents diagnosed with one of the three nationally notifiable sexually transmitted diseases as captured by the CDC's notifiable disease surveillance system between 2012 and 2019.

#### 3.3.2 AtlasPlus

STI diagnosis is based on a set of unified case definitions provided by the Council of State and Territorial Epidemiologists (CSTE) for reporting purposes for each of the nationally notifiable diseases. Confirmed diagnoses of any of the notifiable diseases are sent by physicians, laboratories, or hospitals and clinics to CDC through the Nationally Notifiable Disease Surveillance System (NNDSS) and is later de-identified and made publicly available through AtlasPlus. Case definitions for chlamydia, gonorrhea, and syphilis as defined by CSTE can be found in table A1 of the appendix. Due to the combination of three stages of syphilis infection (primary, secondary, and early non-primary non-secondary) into a single indicator for this analysis all three relevant definitions are presented.

Nationally notifiable surveillance data is compiled by CDC based on reports submitted by health departments from all 50 states, the District of Columbia, and select smaller governing units. Data is available at up to five geographic units: national, regional, state, county, and metropolitan statistical areas covering several health conditions including HIV/AIDS, STIs,

hepatitis and tuberculosis as well as six social determinants of health including insurance status and households living below the poverty level. I used county-level data on STI cases from this data source, including case numbers of chlamydia, gonorrhea, primary and secondary syphilis, and early non-primary non-secondary syphilis. Total syphilis case numbers calculated by adding the three syphilis values together were used for the purposes of analysis.

STI case data for all years between 2012 and 2019 were included in the analysis to capture the period immediately following the conclusion of the NSDUH survey period and avoid possible data complications introduced by the COVID pandemic. Cases counts of chlamydia, gonorrhea, and syphilis were summed across the 2012-2019 period by substate region as well as aggregated into a substate region-wise total STI case count for inclusion in the predictive models.

### 3.3.3 Data Management

To appropriately merge all data at the substate region level, I first created a proprietary crosswalk dataset to assign every US census tract to its state, county, and NSDUH substate region. Census tract data was downloaded from the Census Gazetteer Files:

<https://www.census.gov/geographies/reference-files/2000/geo/gazetter-file.html>. State and county data was taken from the 2000 Census Block Maps available at:

<https://www.census.gov/geographies/reference-maps/2000/geo/2000-census-block-maps.html>.

Substate region codes found in table P8 of the 2002-2011 Codebook were matched to the county or municipal unit categorization found in tables Q1-Q51 also in the Codebook.

Substate region definitions are the product of discussion between SAMHSA and state officials in charge of substance abuse grant money. Definitions were recommended by each state to NSDUH and accepted if adequate sample size was available to create meaningful estimates,



leading to heterogeneity in how substate regions are defined. Most states chose to define substate regions by county with a few (e.g. the District of Columbia, Los Angeles, and the states of Massachusetts and Connecticut) using census tracts or smaller municipal units such as service planning areas or wards to define regions. Due to mismatches in documentation these areas were removed from the analysis. This crosswalk allowed me to merge county-level CDC data with substate-level NSDUH data described previously.

#### 3.3.4 Data analysis

STI data from CDC downloaded as year-specific excel files were brought into Stata and merged before being merged with the NSDUH dataset from Aim #1. Descriptive analyses of the four STI outcomes (chlamydia, gonorrhea, syphilis, and total STI case counts) using the `gladder` command produced a panel of transformations according to the ladder of powers were created for each of the four outcomes to get a preliminary idea of which modeling approach would best fit the data (appendix figure A1). In all four cases a square root or log transformation appeared to provide the best fit. Based on this finding both a generalized linear model (GLM) and negative binomial model were run to determine would be the most appropriate model moving forward. A log transformed version of all four outcomes was created and included in a crude GLM regressing each outcome individually on religiosity. Negative binomial models (NBREG) with case counts similarly regressed on religiosity were run for comparison. The result of the two differing modeling approaches can be seen in appendix table A2. Discrepancy in the results of the syphilis models required further comparison of the NBREG and GLM models using the r-squared values. In all cases the r-squared value was lower for the negative binomial model, leading me to choose NBREG as my preferred model. This is consistent with the nature of the

outcomes as count variables and prevents potential concerns related to altering the data distribution by applying the log transformation.

### 3.3.5 Exploring Alternative Specifications

#### 3.3.5.1 Truncated Negative Binomial Regression

Because I had no regions with zero counts of any STI outcome this feature of the data motivated my checking of my negative binomial model estimates with an alternative specification of 'zero truncated binomial regression' that has been suggested when the count distribution contains no zero count values (Cameron & Trivedi, 2009). The resulting estimates did not differ appreciably from my primary nbreg estimates (Table A3 in the appendix), leading to my decision to continue with the standard negative binomial model.

#### 3.3.5.2 Poisson Regression

I was additionally interested in whether the Poisson modeling approach might yield the same results as the nbreg modeling results. I re-specified my models for the Poisson distribution family and a log link function. An assessment of the chi-squared goodness of fit statistic for the Poisson models revealed that they were highly significant, indicating that the data may be over-disbursed and therefor better modeled by the negative binomial model. An assessment of the mean and variance of each of the STI outcomes showed a large difference between the two values for all outcomes, solidifying the choice of the NBREG model for further data analysis. A table of the mean and variance values can be found in table 4 of the appendix.

### 3.3.6 Crude models

After obtaining the results of the crude unadjusted models of my four STI outcomes generated from the above analysis I sought to add additional basic covariate predictors to my model, starting with population size. Substate regions vary widely in both geographic and

population size. To determine the role that population size may play in predicting STI cases a standardized population term was created by taking the weighted population size of each region provided by NSDUH, subtracting the population mean and dividing by the standard variation.

$$zpop = (WeightedCount - 643148.8)/584445$$

This standardized population term (zpop) was added to the crude models to examine whether population size affected the predictive ability of religiosity.

In addition to population size I was interested in whether the age distribution of the substate region affected STI outcomes. Variable CATAG3 from DAS was brought into the data set which provided the substate region-specific proportion of individuals in five age groups: 12-17, 18-25, 26-35, 36-49, and 50+.

### 3.4 Aim 3

#### 3.4.1 Additional Covariates

To improve the predictive ability of the crude models derived in Aim #2 several additional covariates drawn from the 2002-2011 NSDUH survey were considered for inclusion. Each covariate was added individually to the four outcome models to examine its effect on the religiosity prediction.

Variables with multiple levels were dichotomized within the DAS system to create several covariates. For example, the race variable RACE4 has four levels: White not Hispanic, Black not Hispanic, other or multiple not Hispanic, and Hispanic. To obtain the covariate for Non-Hispanic White the variable was recoded as White not Hispanic and 'otherwise' which included the three remaining categories. This allowed me to separate out the proportion of each substate region who identifies as White not Hispanic from other racial and ethnic identities and then download the data from DAS to be included in the full models. A similar procedure was

followed for all multi-level variables. The included covariates for age are interdependent and take the following form:

1. % age 12-17 years
2. % age 18-25 years
3. % age 26-35 years
4. % age 36-49 years
5. % age 50 years and older

As these five values sum to 100% to avoid a model misspecification, I deliberately omitted one of the covariate terms in these sets where the sum of the percentages was 100%. For example, in the case of the age percentages, I omitted the category of age 50 years and older.

However, in my description of the sample of substate regions I have shown the mean and interquartile range (IQR) for all of these percentages, including the omitted covariate (Table 2).

A table of the covariates with descriptions of how the data was re-coded in the DAS system (where applicable) can be found in table A5 of the appendix.

**Table 2.** Mean and interquartile range for all covariates considered in my modeling approaches for Aim #3

Covariate	Mean	IQR
Population	0.003	-0.64, 0.38
% age 12-17	0.10	0.09, 0.11
% age 18-25	0.13	0.12, 0.15
% age 26-35	0.14	0.12, 0.15
% age 36-49	0.25	0.24, 0.27
% age 50 up	0.37	0.34, 0.41
Non-Hispanic White	0.76	0.66, 0.91
Hispanic ethnicity	0.09	0.02, 0.10
Gender	0.48	0.47, 0.50
AUD	0.08	0.06, 0.09
DUD	0.03	0.02, 0.03
Poverty	0.12	0.09, 0.15
Education	0.14	0.11, 0.17
Employment	0.48	0.44, 0.51
Insurance coverage	0.14	0.10, 0.16

**Table 2 (cont'd.)**

Booked for a crime	0.17	0.15, 0.19
Tobacco usage	0.31	0.28, 0.34
Extra-medical drug use	0.06	0.05, 0.07

#### 3.4.1.1 Extra-medical drug use definition

Extra-medical drug use refers to the usage of drugs falling into four categories: analgesics, stimulants, tranquilizers, and sedatives, in a context in which the drug was not prescribed or it was taken only for the feeling it causes. These drugs are referred to as ‘psychotherapeutic drugs’ by NSDUH, however, because the category does not include the use of antipsychotic medications and is not concerned with prescribed usage it will be referred to as ‘extra-medical drug use’ in this study (Parker & Anthony, 2015).

#### 3.4.2 Modeling approach

Covariates were initially assessed individually or in thematically appropriate groups (e.g. age, race/ethnicity). Following the assessment of covariates individually on religiosity for each of the four outcomes, three modeling iterations were used to arrive at the final predictive model. Covariates identified as probable predictors of STI outcomes were the first to be added to the crude models generated in Aim #2. Statistically significant predictors at the  $p < 0.05$  level from the probable model were retained for the next iteration in which unlikely predictors were additionally added. Predictors for each outcome were again retained and run in the third and final ‘best fit’ model.

#### 3.4.3 Post-Estimation Exploratory Data Analysis

After fitting all four final models I created plots of residual versus fitted (RVF) values for each final model as well as the plausible predictor models and unlikely predictor models. To generate the plots Stata requires the use of the GLM command as opposed to the NBREG command I used for modeling up to this point. GLM allows for the specification of family and

link meaning I was able to again specify the use of the negative binomial distribution and a log link, however the resulting model does not fully align with the output from NBREG.

Notwithstanding, I created three RVF plots for each outcome, one for the initial probable predictor model, a second for the unlikely predictor model, and a third for the final model.

## Chapter 4. Results

### 4.1 Aim 1

#### 4.1.1 Descriptive statistics

The 2002-2011 NSDUH restricted data set automatically weights all variables to protect the privacy of respondents. Table 3 provides the weighted characteristics of the sample as derived from the NSDUH restricted use data system (DAS).

**Table 3.** Descriptive statistics of the NSDUH data set used for this research

Gender	%	n
Male	48.5	119,423,000
Female	51.5	126,905,000
<b>Race/Ethnicity</b>		
Non-Hispanic White	68.3%	168,210,000
Non-Hispanic Black	11.6%	28,636,000
Non-Hispanic Native American / Alaskan Native	0.6%	1,454,000
Non-Hispanic Native Hawaiian / Pacific Islander	0.3%	770,000
Non-Hispanic Asian	4.3%	10,597,000
Non-Hispanic Mixed Race	1.2%	2,963,000
Hispanic	13.7%	33,697,000
<b>Age</b>		
12-17	10.1%	24,977,000
18-25	13.3%	32,780,000
26-34	14.4%	35,515,000
35-49	26.1%	64,172
50+	36.1%	88,883,000
<b>Religion</b>		
<b>Decision</b>		
Strongly Agree / Agree	65.1%	160,431,000
Strongly Disagree / Disagree	23.7%	58,400,000
Otherwise / Youth	11.2%	27,497,000
<b>Importance</b>		
Strongly Agree / Agree	67.5%	166,392,000

**Table 3 (cont'd.)**

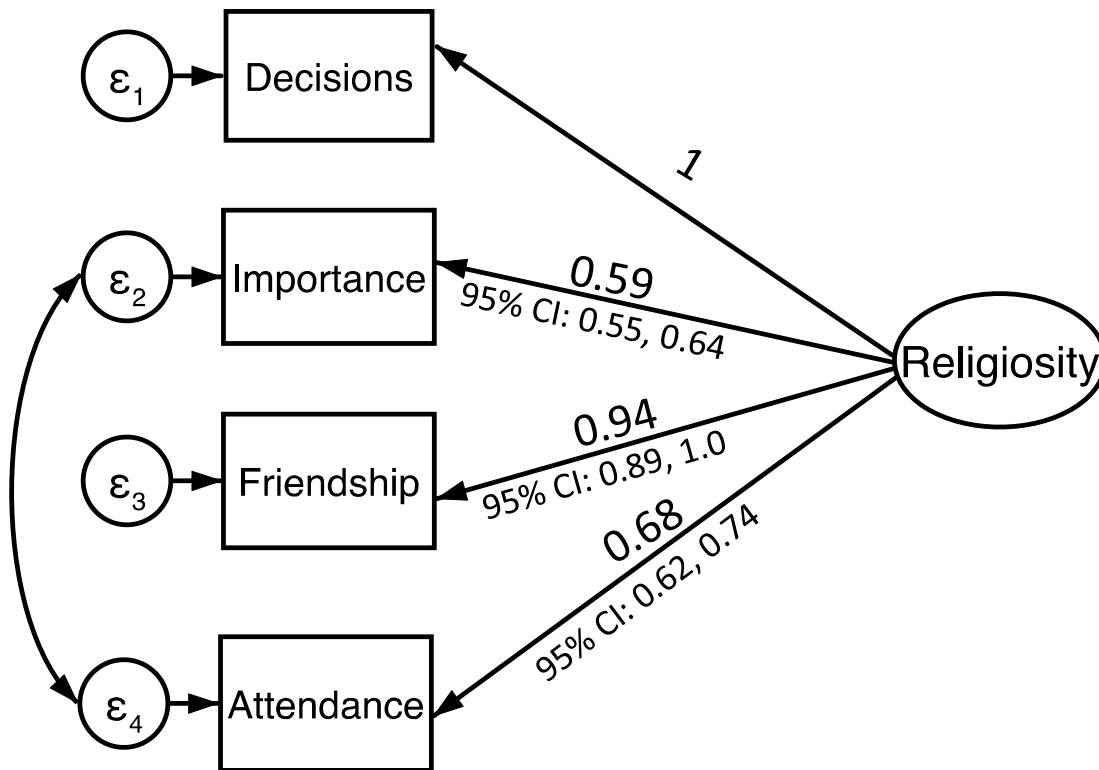
Strongly Disagree / Disagree	21.2%	52,322,000
Otherwise / Youth	11.2%	27,614,000
Friendship		
Strongly Agree / Agree	29.8%	73,393,000
Strongly Disagree / Disagree	58.9%	145,074,000
Otherwise / Youth	11.3%	27,861,000
Attendance		
0	31.6%	77,954,000
1-2	9.6%	23,719,000
3-5	8.9%	21,805,000
6-24	12.4%	30,516,000
25-52	14.0%	34,572,000
52+	12.8%	31,552,000
Weighted Total		
	100%	246,328,000

#### 4.1.2 Latent Trait Analysis

I fit a CFA model to the item-level religion data from NSDUH. The CFA confirmed my hypothesis of the existence of a single latent dimension which I termed ‘religiosity’. However, upon examining the modification indices for the CFA it appeared that there might be a violation of the local independence assumption. The local independence assumption states that there is no correlation between residual terms, as there appeared to be some correlation present I re-specified the 1-dimensional model with correlation between the residuals for the importance and attendance items (Figure 1). The result indicated that the re-specification satisfied the local independence assumption.

Based on the resulting model, I used a post-estimation prediction to derive an estimated religiosity score for each substate region, which became my main predictive covariate in this dissertation research project. This procedure allowed me to use a two-step process. First, in the final estimation step of Aim 1 I derived the estimated position of each substate region on the dimension of religiosity. I then used that predicted score as my main predictor for Aims 2 and 3.





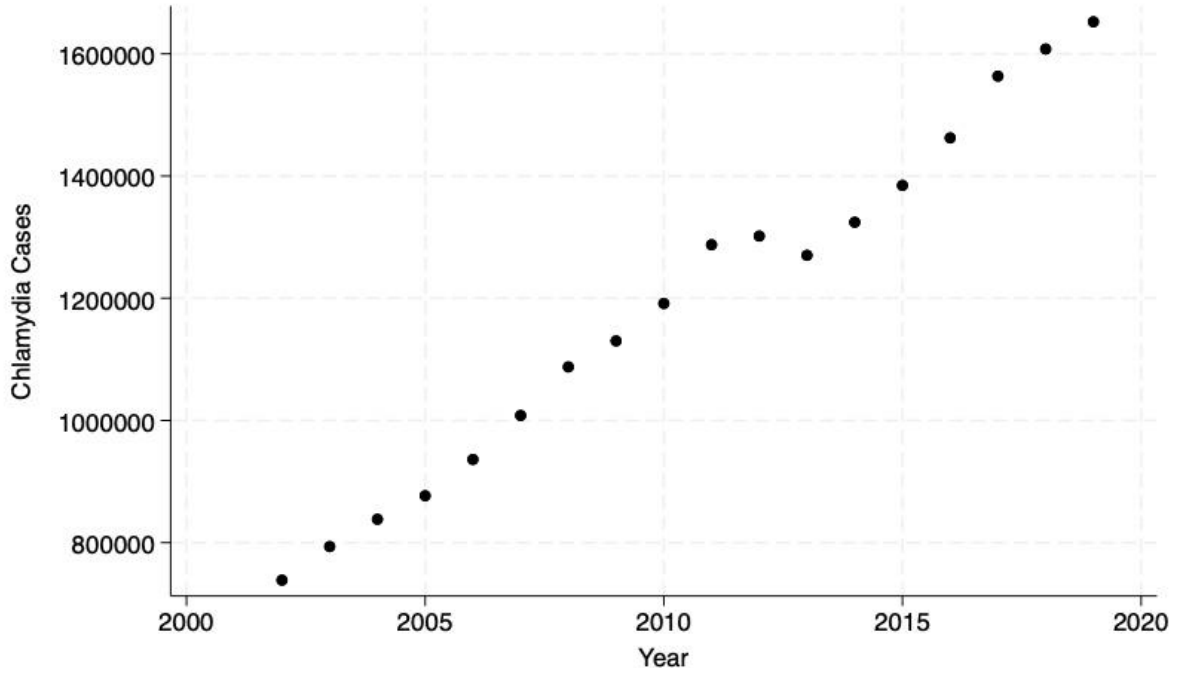
**Figure 1.** CFA used to predict religiosity from the four NSDUH religion items. Items are indicated by rectangles while the latent dimension is indicated by an oval. Arrows from the latent trait to observed variables show the factor loadings and respective 95% confidence intervals for each variable.

## 4.2 Aim 2

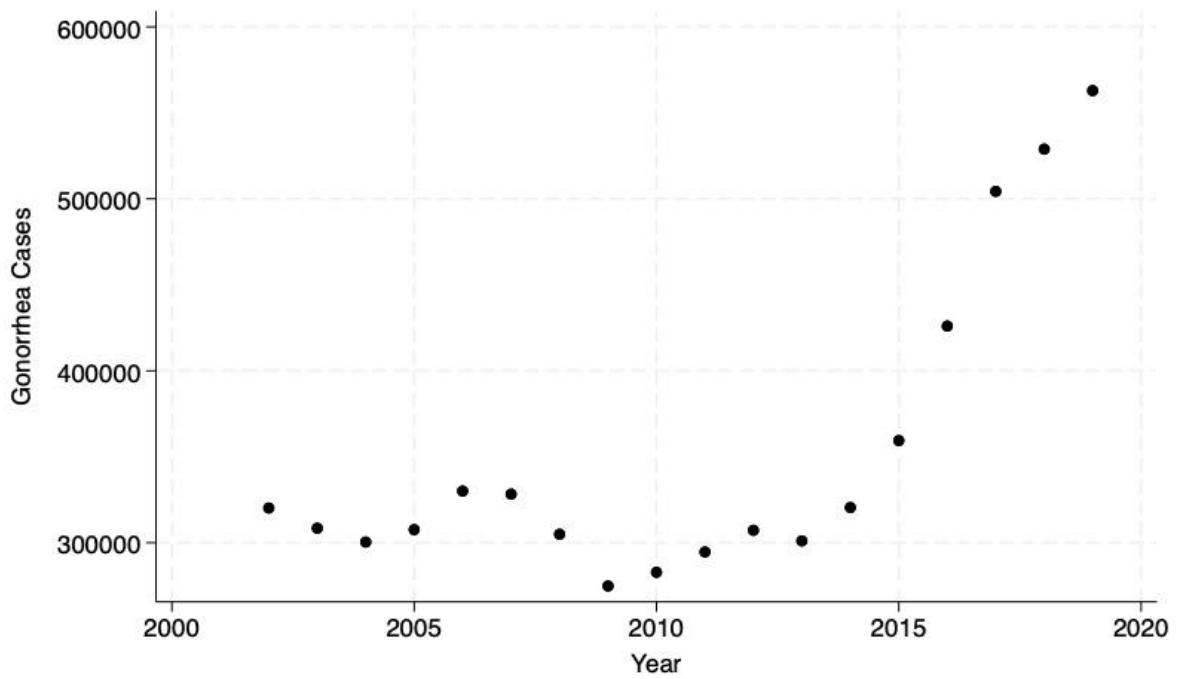
### 4.2.1 Descriptive Statistics

Although there were 383 total substate regions defined by NSDUH for the 2002-2011 surveys, this analysis included only 354 substate regions. As previously described, use of census tracts to define regions in some states led to exclusion of those regions from analysis. Areas excluded from analysis include all of Connecticut and Massachusetts, two regions in Delaware, one in Michigan, the city of Los Angeles, and the District of Columbia. Case numbers of chlamydia, gonorrhea, and syphilis for the 354 included substate regions from 2002 to 2019 are

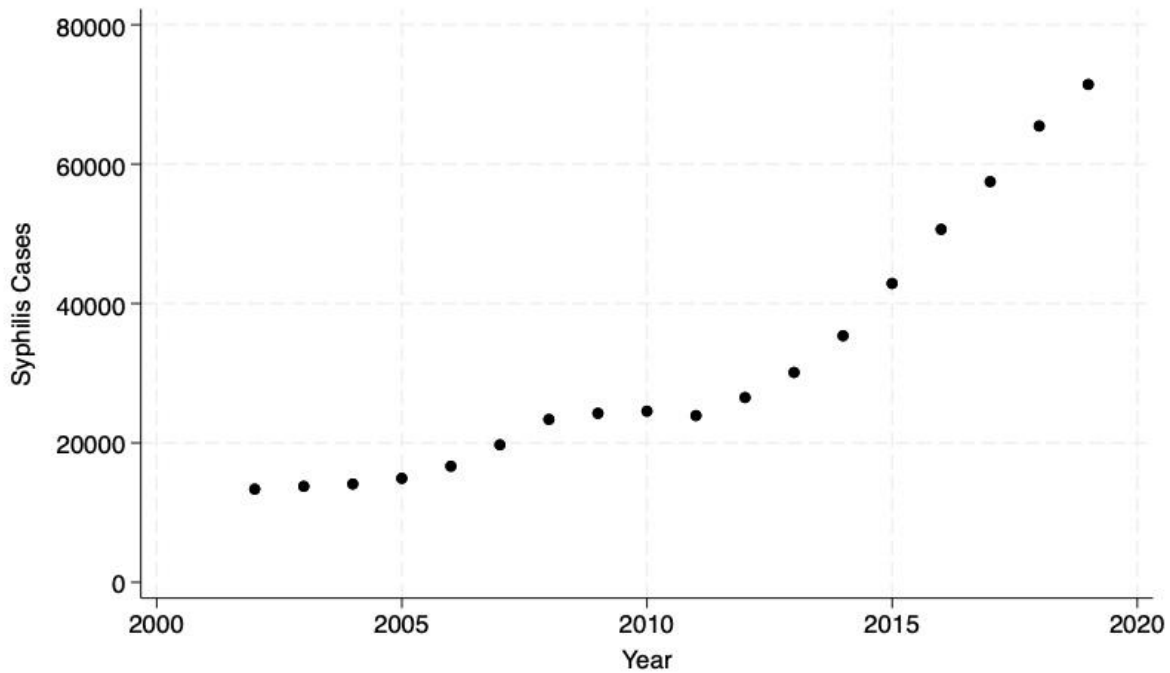
presented in figures 2-4. Between 2002 and 2019 case numbers of all three STIs increased substantially in the US.



**Figure 2.** Chlamydia cases in the US reported to CDC between 2002 and 2019.



**Figure 3.** Gonorrhea cases in the US reported to CDC between 2002 and 2019.



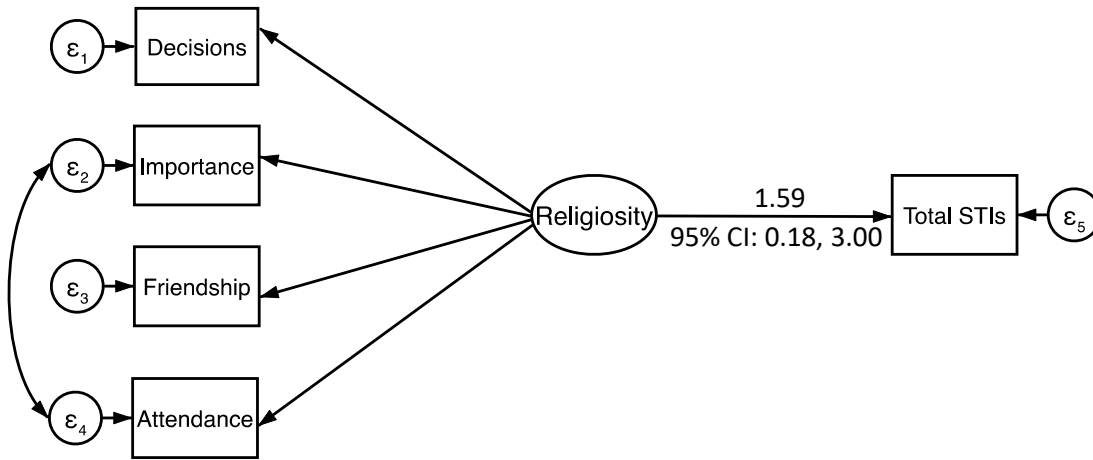
**Figure 4.** Primary, secondary, and early non-primary non-secondary syphilis cases in the US reported to CDC between 2002 and 2019.

#### 4.2.2 Approach – Crude Models

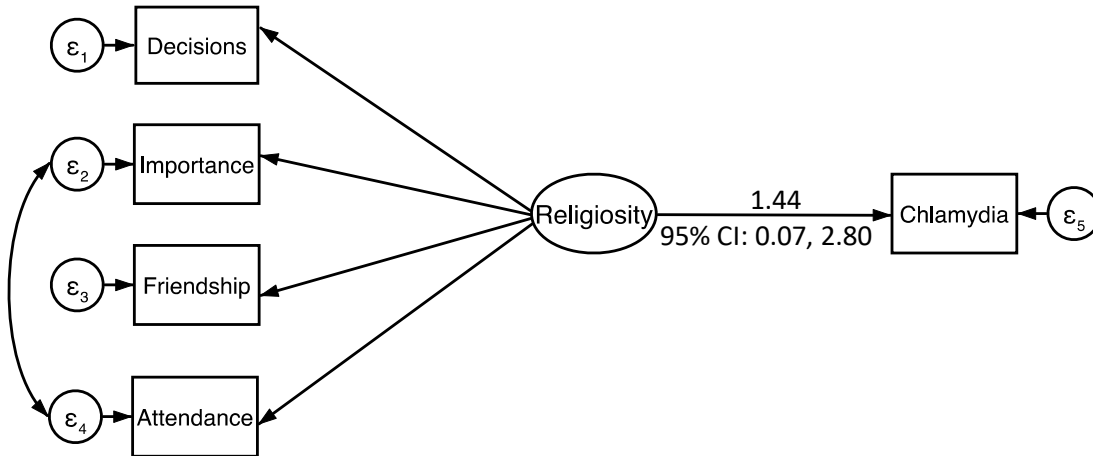
As specified for Aim 2, I began the modeling process by specifying a covariate-unadjusted crude negative binomial model regressing total substate region STI cases on the substate-level predictor of religiosity derived from the 2002-2011 NSDUH data in Aim 1. The results of the model showed that religiosity was predictive for predicting total STI cases by religiosity at the substate level.

Because CDC provides STI counts I first fit a Poisson model, however I discovered over-dispersion and turned to a negative binomial model (appendix table A4). I then repeated the estimation steps for each of the three nationally notifiable STIs individually. Figures 5-8 provide a cartoon representation of the resulting models with the estimated slopes and associated 95% confidence intervals. Finally, because substate region-specific counts are influenced, in part, by the population of the region, I re-fit the models to include a covariate term for region-specific

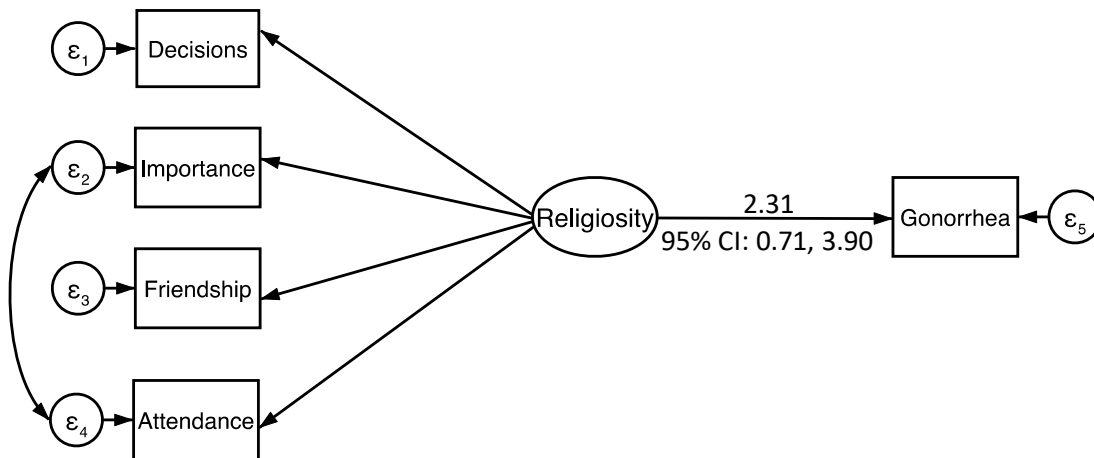
population size (appendix figure A2).



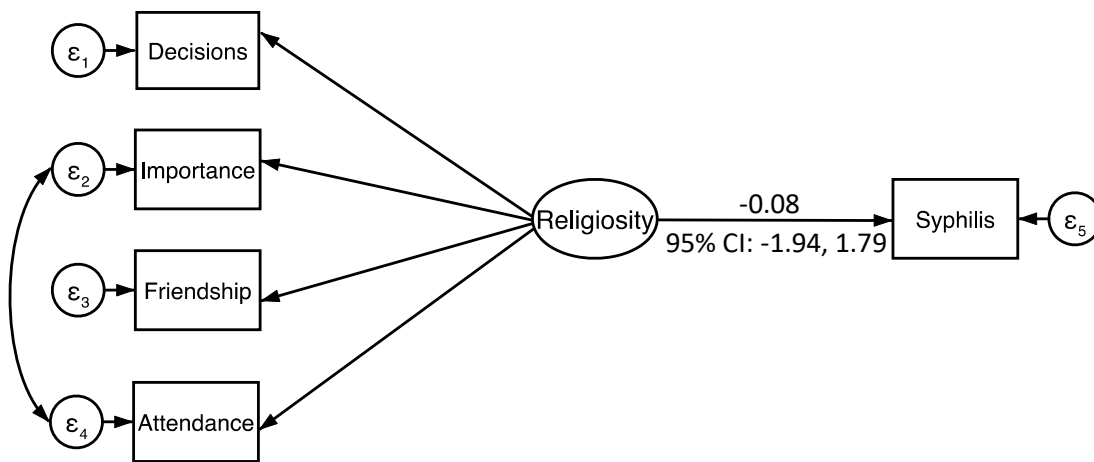
**Figure 5.** Cartoon depiction of the crude unadjusted model of total STI cases with estimated slope and 95% confidence interval.



**Figure 6.** Cartoon depiction of the crude unadjusted model of chlamydia cases with estimated slope and 95% confidence interval.



**Figure 7.** Cartoon depiction of the crude unadjusted model of gonorrhea cases with estimated slope and 95% confidence interval.



**Figure 8.** Cartoon depiction of the crude unadjusted model of syphilis cases with estimated slope and 95% confidence interval.

When total STI cases are broken into the three nationally notifiable diseases rather than viewed in aggregate differing patterns in the predictive capacity of religiosity emerge. Chlamydia and gonorrhea infections follow a similar pattern to overall cases, increasing along with substate region-level religiosity. Syphilis, however, appears to decrease as religiosity increases but is not predicted by religiosity in the crude model. When a standardized population

term is included it increases the strength of religiosity as a predictor of all four outcomes. This is especially notable in the case of syphilis where the association between religiosity and log syphilis cases at the substate region level becomes stronger than it was in the crude model.

#### 4.3 The Issue of Age

After observing the effect of the addition of the population term on the models the next step was to investigate whether the age distribution of the substate regions was predictive of STI outcomes and whether age also affected the predictive capacity of religiosity. Using age group 50+ as the reference category, four age variables representing the proportion of the substate region population in each age group (12-17, 18-25, 26-35, 36-49) were added to the previous model. In all four models where age distribution was included religiosity remained an important predictor of the outcome as well as population size. However, not all the age group proportions were predictive of STI outcomes. Table 4 below displays the results of the addition of covariates for population and age for each of the four STI outcomes.

**Table 4.** Results of the addition of covariates for population size and age to each of the four STI models. LR test and p-values compare each model to the crude unadjusted model

Total STI model +	Religiosity slope (95% CI)	Improvement to religiosity model - LR test	Model p > chi <sup>2</sup>
Religiosity	1.59 (0.18, 3.00)	--	--
Religiosity + population	2.35 (1.52, 3.17)	359.77	<0.001
Religiosity + population + age 12-17	2.23 (1.46, 3.18)	0.06	0.8140
Religiosity + population + age 12-17 + age 18-25	2.22 (1.39, 3.06)	22.92	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35	2.38 (1.58, 3.17)	54.50	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35 + age 36-49	2.45 (1.63, 3.27)	55.03	<0.001
Chlamydia model +			
Religiosity	1.44 (0.07, 2.80)	--	--
Religiosity + population	2.12 (1.34, 2.89)	381.50	<0.001

**Table 4 (cont'd.)**

Religiosity + population + age 12-17	2.01 (1.21, 2.81)	0.95	0.3298
Religiosity + population + age 12-17 + age 18-25	1.91 (1.14, 2.69)	21.88	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35	2.02 (1.27, 2.77)	48.67	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35 + age 36-49	2.08 (1.31, 2.85)	49.06	<0.001
<b>Gonorrhea model +</b>			
Religiosity	2.31 (0.72, 3.90)	--	--
Religiosity + population	3.41 (2.34, 4.49)	267.63	<0.001
Religiosity + population + age 12-17	3.68 (2.53, 4.83)	1.87	0.17
Religiosity + population + age 12-17 + age 18-25	3.68 (2.57, 4.79)	26.05	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35	4.01 (2.94, 5.08)	60.16	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35 + age 36-49	4.11 (3.01, 5.21)	60.78	<0.001
<b>Syphilis model +</b>			
Religiosity	-0.08 (-1.94, 1.79)	--	--
Religiosity + population	1.02 (-0.12, 2.16)	287.15	<0.001
Religiosity + population + age 12-17	1.70 (0.43, 2.98)	6.11	0.01
Religiosity + population + age 12-17 + age 18-25	1.57 (0.36, 2.79)	36.86	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35	2.12 (0.97, 3.27)	93.18	<0.001
Religiosity + population + age 12-17 + age 18-25 + age 26-35 + age 36-49	2.20 (1.01, 3.39)	93.50	<0.001

#### 4.4 Effect of Individual Predictors on Religiosity

The addition of covariates individually to the population-adjusted outcome models revealed that religiosity remained predictive of log total STI cases at the substate region level in all cases (Table 5). Non-Hispanic White, poverty status, educational attainment, and insurance coverage reduced the predictive ability of religiosity on the outcome while alcohol and drug use

disorders increased the importance of religiosity. The results of individual covariate addition to the individual STI outcome models can be found in appendix tables A6-A8.

**Table 5.** Effect of individual covariate addition to the population-adjusted crude model on the slope of the religiosity – total STI outcome relationship.

Total STI model +	Religiosity slope (95% CI)	Improvement to religiosity model - LR test	Model p > chi <sup>2</sup>
Religiosity	1.59 (0.18 – 3.00)	--	--
Religiosity + population	2.35 (1.52 – 3.17)	359.77	<0.001
Religiosity + population + age	2.45 (1.63 – 3.27)	55.03	<0.001
Religiosity + population + race/ethnicity	1.22 (0.46 – 1.98)	113.71	<0.001
Religiosity + population + gender	2.11 (1.29 – 2.93)	19.03	<0.001
Religiosity + population + AUD	3.23 (2.31 – 4.16)	13.72	0.0002
Religiosity + population + DUD	3.52 (2.72 – 4.32)	76.75	<0.001
Religiosity + population + poverty	1.09 (0.16 – 2.01)	29.82	<0.001
Religiosity + population + education	1.54 (0.55 – 2.52)	8.56	0.0034
Religiosity + population + employment	2.36 (1.52 – 3.19)	0.04	0.8325
Religiosity + population + insurance coverage	1.92 (0.99 – 2.85)	3.85	0.0489
Religiosity + population + booked for a crime	2.33 (1.51 – 3.15)	2.70	0.1004
Religiosity + population + tobacco usage	2.46 (1.59 – 3.34)	0.70	0.4034
Religiosity + population + extra-medical drug use	2.82 (1.99 – 3.64)	28.41	<0.001

#### 4.5 Plausible Predictors

Beyond the age distribution and population size there are several covariates available from NSDUH that I was interested in exploring. In this step I included the most plausible predictors from my list of covariates including: Non-Hispanic White, Hispanic ethnicity, alcohol use disorder, drug use disorder, poverty status, gender, and educational attainment. These



predictors are covariates which have been previously shown in the literature to influence STI outcomes. The results of the negative binomial model showed that seven of the variables were predictive of total substate region STI cases. Three covariates: male gender, Non-Hispanic White, and Hispanic ethnicity were associated with decreased STI cases in the substate regions. The remaining four: religiosity, percentage of the population ages 26-35, population size, and drug abuse were predictors of increased total STI cases. The results of these initial models can be found in appendix tables A9-A12.

Like the crude models, both the expanded chlamydia and gonorrhea outcome models returned the same predictors as the total substate STI model. However, unlike the other three outcomes, syphilis was not predicted by religiosity or male gender and the percentage of youth aged 12-17 in the population was a negative predictor of substate syphilis cases.

#### 4.6 Unexpected Predictors

After specification of the models to include my plausible predictors I created a second round of further expanded models, retaining the significant predictors from the previous models, and adding additional covariates that I was less sure would be predictive of STI outcomes. The addition of the second group of covariates led to a greater number of differences between the predictors of each outcome than in previous steps. Several covariates were or remained important in all four models: Hispanic ethnicity, Non-Hispanic White, proportion aged 26-35, and population size. For the chlamydia and total STI models additional predictors included: religiosity, male gender, being fully employed, being uninsured, and extra-medical drug use. In contrast to chlamydia, male gender and being fully employed were not predictive of substate gonorrhea cases, but tobacco use was. Finally, in the model of the syphilis outcome the

proportion aged 12-17 and ever having been booked for a crime were present, both predictors were unique to the syphilis model (results in tables A12-A15 of the appendix).

#### 4.7 Summary Model and Sensitivity Analyses

Based on the results of the series of models created for each outcome in previous steps I created a best fit model for each outcome. To prevent a potential table 2 fallacy the results of all four models are not presented but are available upon request to the author. There was considerable overlap between all the models; Hispanic ethnicity, Non-Hispanic White, proportion of the substate region population ages 26-35, and extra-medical drug use were universally predictive of the STI outcomes under study.

#### 4.8 Post-Estimation Exploratory Data Analysis

Comparing the outcome of successive models I expect to see improvement to the fit of the model. In the case of all four outcomes there were improvements in fit, however in all cases several outliers heavily skewed my data. These outliers contributed to a steep negative slope of the points on the right-hand side of the distribution for all outcomes (plots available in appendix figure A3). Because the RVF plots showed that the data was heavily skewed I will explore whether a multivariable fractional polynomial (MFP) approach is a better fit for my data as part of the post-estimation exploratory data analysis. While my study has assumed the presence of a single slope the MFP procedure relaxes the assumption of a single slope and may allow me to capture residual variation.

## Chapter 5. Discussion/Limitations

The main findings of this study may be summarized succinctly. First, the four religion questions asked by the NSDUH survey can be reliably used to measure latent religiosity at the substate region-level. Second, substate-level religiosity alone is a predictor of chlamydia, gonorrhea, and total STI outcomes at the substate level. Only in the case of syphilis is it not predictive. Perhaps surprisingly, religiosity exhibits a positive relationship with STI outcomes. Third, while there were some universally important predictors of STI outcomes there was heterogeneity in the covariates which increased the predictive capacity of the models.

One of the strengths of this project is the use of publicly available data as well as the linkage of two large datasets. All the code used for this project including the crosswalk file have been made publicly available so that future work can build upon the methods and findings of this study.

### 5.1 Aim 1

To my knowledge, this is one of few if not the only study to use latent trait analysis to examine religiosity from the NSDUH survey. Previous studies assessing religiosity using the NSDUH survey have included some or all the religion variables in modeling but have included them as individual covariates rather than as a single measure of a latent religiosity trait as was done here. This work thus extends previous work done using the NSDUH religion variables to measure religiosity and shows that the survey questions included in the NSDUH survey reliably predict underlying religiosity.

#### 5.1.1 Limitations

This approach has several limitations: 1. NSDUH survey responses are self-reported and thus potentially subject to social desirability bias. 2. The four religion variables collected by

NSDUH are largely focused on the performance of religion. The results of this study may therefore not be replicable if conducted using an alternative set of questions to measure religiosity. 3. The NSDUH restricted use data is not available at the individual level and despite being epidemiologically representative samples of the US population they are cross-sectional, preventing this study from drawing causal conclusions. 4. A factor score is an imperfect predictor of the values of a latent trait. Using the factor score simplifies the specified models, however, it introduces some measurement bias as it is an estimate of the latent variable based on the mean of the latent variable conditional on the observed variables in the model. 5. This analysis includes only responses from individuals 18 and older. While NSDUH collects responses on the four religion variables from youth aged 12-17 the SEM failed to achieve convergence despite a factor analysis showing strong loading of all four variables to a single factor as with the adult responses.

Notwithstanding these limitations these results are of interest as they show that substate level religiosity can be predicted by the four NSDUH religion variables. This is an important proof of concept for future studies which can build on the methods used here to improve the measurement of religiosity. Repeating this study utilizing the latent trait as opposed to the factor score during modeling to compare with the results presented here is an important next step for future investigation.

## 5.2 Aim 2

These results show that substate region-level religiosity is predictive of chlamydia, gonorrhea, and total STIs but not syphilis. Perhaps surprisingly the relationship between religiosity and STI outcomes is strongly positive, indicating that more highly religious substate regions are more likely to have higher numbers of chlamydia, gonorrhea, and total STI cases.

Accounting for heterogeneity in the size of substate region populations by adding a standardized population covariate into the model showed that population size alone was a predictor of STI outcomes, however it also improved the predictive ability of religiosity for all four STI outcomes. The importance of population size as a predictor of STI outcomes makes intuitive sense as larger populations have a greater number of people to become potentially infected. The improvement to the predictive ability of religiosity when the population term was added indicates that regardless of population size religiosity is a strong predictor of STI outcomes, strengthening the findings of the crude model. This was further strengthened by the addition of age terms into the model. The finding that the proportion of individuals aged 18-25 and 26-35 in a substate region was predictive of three of the four outcomes matches national data which has shown that these groups bear the overwhelming majority of the STI burden in the US (Kreisel et al., 2021).

To my knowledge, this is the first study to join NSDUH and CDC data to examine religiosity and sexually transmitted infections in the US. There is a large and growing body of research regarding the relationship between religion and STIs but little agreement on whether religiosity plays a protective or

**Sidebar 1.** Syphilis is a communicable disease caused by infection with a Gram-negative diderm bacteria *Treponema pallidum* subsp. *pallidum*. Transmission of syphilis can occur when an uninfected host comes into contact with infectious lesions called chancres, often via sexual activity and sometimes with transmission from a pregnant mother to the fetus. Bacteria are generally shed from the lesion as mobile flagellate spirochetes.

According to Peeling and colleagues, effective transmission between an infected individual and a susceptible host may require as few as ten spirochetes (2017). Small tears in the anal and genital regions during sex are common due to the thinness of the skin in these areas with subsequent increased risk of infection. Once the infection occurs, these secondary cases seem to be most likely to transmit the infection and disease to others while they are in the early stages of the disease, generally up to 2 years post-infection.

In contrast to syphilis, chlamydia and gonorrhea are transmitted through contact with infected genital fluids. When used correctly, male condoms can lower the risk of chlamydia and gonorrhea by up to 90% (Marfatia et al., 2015). In the case of syphilis, the condom approach can reduce the risk of infection, but protection is complete only when the condom fully cover the chancre or other lesion.

The variations in the infection-specific transmission and the relative protectiveness of condoms might help explain the project's observed variations in its

negative role in STI outcomes. This research adds to the literature exploring the role of religiosity in sexually transmitted infection outcomes in the US and confirms the importance of 18–35-year-olds as a high-risk population.

### 5.2.1 Limitations

Limitations of this aim include: 1. NSDUH and CDC report data at differing geographic units, leading to some mismatch between the data sets and elimination of some regions from analysis. I will conduct post-estimation exploratory data analysis to examine potential demographic differences between the included and excluded regions. 2. All three of the STIs included in this study can be asymptomatic, there are likely hundreds of thousands of asymptomatic infections that are not included in the CDC case counts. 3. CDC data comes from notifiable disease reports sent in by physicians and health organizations, reporting may be incomplete due to delays or failures to report cases or failure to seek care for asymptomatic or mild infections.

The asymptomatic nature of the three STIs discussed here is an important limitation of this study. Chlamydia, the most common of the three reportable STIs is also the most likely to be a silent infection. It is estimated to be asymptomatic in as many as 77% of cases (Farley, Cohen, & Elkins, 2003). While the CDC reported over 1.64 million cases in the US in 2021 the true number of cases is likely at least double that figure (CDC, 2021). Asymptomatic infections lead to chlamydia going undiagnosed or only being diagnosed when medical care is sought for another STI or gynecological condition, limiting the accuracy of the case reports provided to and by CDC. However, CDC notifiable disease reports are the most comprehensive source of STI data available and thus were used for this study.

### 5.2.2 Public Health Implications

This study adds to the existing public health literature on the role of religion in sexual health and confirms the importance of young adults in driving STI outcomes in the US. Confirming the importance of religiosity as a predictor of STI outcomes supports the need to account for this dimension when designing public health programs in the future. Public health programs which are able to tailor their outreach to more religious communities as well as younger individuals may have a greater impact on reducing STI cases than those which ignore the role that religiosity plays in STI outcomes.

### 5.3 Aim 3

In this aim I built upon the models in Aim #2 to assess the importance of additional covariates to predicting STI prevalence at the substate level and present evidence of heterogeneity in substate region level predictors of STI outcomes. Taken together these analyses provide some evidence of differences in risk factors for sexually transmitted infections. My hope is that this work can help guide further studies to eventually inform public health outreach, intervention, and resource allocation intended to reduce the burden of STIs in the US population.

The focus of this dissertation was on the possibility that a region's level of religiosity might belong in the conceptual models we use to predict the occurrence of sexually transmitted infections and possibly to prevent and control these STIs. The analysis steps of the dissertation identified a set of other potentially important predictive covariates as listed here for each specific STI subtype and for all three STI subtypes considered together. For chlamydia, the predictive covariates that qualified for inclusion in the model were: religiosity, population, age 26-35, Hispanic ethnicity, Non-Hispanic White, proportion male, employment status, insurance coverage, and extra-medical drug use. For gonorrhea, the predictive covariates were: religiosity,

population, age 26-35, Hispanic ethnicity, Non-Hispanic White, insurance coverage, past 30 day tobacco usage, and extra-medical drug use. For syphilis, the predictive covariates were: population, Hispanic ethnicity, Non-Hispanic White, age 12-17, age 26-35, ever booked for a crime, and extra-medical drug usage. For all three STIs considered in aggregate, the predictive covariates were: religiosity, population, age 26-35, Hispanic ethnicity, Non-Hispanic White, proportion male, employment status, insurance coverage, and extra-medical drug use. In order to avoid what now is called the 'Table 2 fallacy' I have not presented the slope estimates for each of these predictive covariates in my results section. The estimates are available upon request to me and will be found in an online repository soon after this dissertation is published. However, I draw attention to the Table 2 fallacy and the difficulty that can be faced when trying to interpret the meaning of the slope estimates in a modeling task of this type (Westreich & Greenland, 2013).

### 5.3.1 Limitations

The limitations of this study are the same as those described above in aim 2 but also include: 1. The variables available for inclusion in this aim were limited to those with sufficient cell size at the substate level so as not to be suppressed by the DAS system. 2. The NSDUH survey does not include information on the specific religious practice of respondents, making it impossible to parse out potential differences due to religious affiliation. 3. The NBREG command does not allow for the inclusion of interaction terms, potentially obfuscating the relationship between predictors.

The lack of interaction terms in the model is an important limitation of this project. To check potential covariance between covariates I ran a series of stepwise backwards selection models with all covariates included in the model. Stepwise regression provides somewhat of a



check of covariance as closely related variables will not remain in the model together due to the small improvement to fit provided by the addition of a closely related variable. The stepwise regression models were specified to allow entry at a significance level of 0.05 and removal at significance of 0.1. Each of the four models returned all the same predictors for their respective outcomes except for the addition of male sex to the syphilis model, supporting the final models.

Additionally important to address is the role religious affiliation may play in predicting STI outcomes as well as affecting the predictive ability of several covariates. Several religions including Islam and Mormonism prohibit the use of alcohol, for example. Areas like Utah with a large Mormon population may introduce unmeasured bias into the model. Additionally, some religious groups have much more strict views of premarital intercourse than others. Some orthodox Jews for example practice ‘shomer negiah’ or a prohibition on touching members of the opposite sex outside of a spouse or immediate family (Charendoff, 2019). On the other hand the unitarian universalist church includes comprehensive coed sexual education as part of their Sunday school curriculum (<https://www.uua.org/re/owl/faq>). With no data on the religious breakdown of the substate regions it is impossible to account for error introduced by varying proportions of adherents to these varied belief systems.

### 5.3.2 Public Health Implications

This dissertation research project has identified a number of covariates as important predictors of STI outcomes at the substate level in the US. The results of the predictive modeling done here points to several subgroups of the population to whom sexual health and STI prevention programming could be targeted to potentially lower STI case numbers in the US. This modeling approach may also be useful in assessing the effect of public health programs. Comparing STI outcomes in substate regions which have implemented new sexual health

programs to similar regions which have not had any alteration of their programs may be an effective method of assessing whether an intervention is effective.

## Chapter 6. Future Directions

There are several extensions of this research that are important future directions. I have shown that it is possible to predict STI outcomes by religiosity and that there are several additional important predictors that can be drawn from the NSDUH survey data. This study used survey data from 2002-2011 to predict STI outcomes but additional data sets are available that could be used to assess whether the findings from this study are applicable to larger samples. For example, a data set encompassing 2002-2017 exists which could be used to predict out to 2018 and 2019 outcomes. Additionally, several smaller datasets exist within DAS with two- and four-year estimates. These datasets could be used individually to predict later STI values then summarized using meta-analysis techniques to allow for overlap of the estimates and demonstrate reproducibility.

Additionally, examining a wider range of STIs would provide greater insight into the drivers of these infections in the US population. According to the CDC three STIs cause much greater burdens of disease in the US population than the three notifiable conditions assessed here (Kreisel et al., 2021). Those three infections are: HPV, herpes, and trichomoniasis which all account for huge numbers of cases and should be included in future studies to determine whether they follow similar patterns to the notifiable infections or if alternative strategies are needed to reduce case numbers.

One of the most important limitations of this study and an important direction for future research is to assess a wider range of covariates to predict STI outcomes. I have not thoroughly assessed whether levels of religiosity differ by ethnic subgroups within the US and whether these potential differences may lead to different predictive values. In part, my hesitation stems from the fact that the NSDUH does not measure respondent's religious affiliation (e.g., Catholic,

Anglican, Greek Orthodox, Islam, Jewish, evangelical Protestant, etc.). Therefore, many of the underlying issues might remain uncertain without these affiliations or religious identities in the dataset for predictive estimates. The interaction between ethnic self-identification and patterns of religious affiliation (e.g., the Church of the Latter Day Saints and non-Hispanic Whites as an example of a potential pattern of specific interest) deserves further consideration with data sets that have more complete coverage of the religiosity dimension. I have had to work with under an assumption of no violation of the measurement equivalence assumption. I have assumed in this dissertation that participants from various subgroups answer the survey items in a similar fashion, making it possible to specify a uniform measurement model, as described under Aim 1. In a more psychometrically oriented dissertation research project, this assumption of a uniform measurement model might have been challenged and this must be faced as a future direction in this line of research.

In addition to the limitations imposed by the lack of data on religious affiliation, this study is limited by the inability to take into account the mode of transmission for any of the STI outcomes. How an individual becomes infected, whether it be due to personal risky behavior or if a partner has exposed them to infection (e.g. through an affair partner) would likely affect what factors are predictive of infection. This type of data is not publicly available from CDC however the method of transmission deserves further consideration using data sets with more complete information on STI diagnoses. Examining the potential differences in predictive estimates of religiosity between sub-groups of method of transmission is an important future direction for research.

## REFERENCES

- Adams, G. C., Wrath, A. J., Le, T., Adams, S., De Souza, D., Baetz, M., & Koenig, H. G. (2020). Exploring the Impact of Religion and Spirituality on Mental Health and Coping in a Group of Canadian Psychiatric Outpatients. *The Journal of Nervous and Mental Disease*, 208(12), 918–924. <https://doi.org/10.1097/NMD.0000000000001243>
- Agli, O., Bailly, N., & Ferrand, C. (2015). Spirituality and religion in older adults with dementia: A systematic review. *International Psychogeriatrics*, 27(5), 715–725. <https://doi.org/10.1017/S1041610214001665>
- Awaworyi Churchill, S., Appau, S., & Ocloo, J. E. (2021). Religion and the Risks of Sexually Transmissible Infections: Evidence from Britain. *Journal of Religion and Health*, 60(3), 1613–1629. <https://doi.org/10.1007/s10943-021-01239-0>
- Baumard, N., & Boyer, P. (2013). Explaining moral religions. *Trends in cognitive sciences*, 17(6), 272–280. <https://doi.org/10.1016/j.tics.2013.04.003>
- Cameron, A.C. and Trivedi, P.K. (2009) *Microeconometrics Using Stata*. Stata Press, Texas.
- Camino-Gaztambide, R. F., Fortuna, L. R., & Stuber, M. L. (2022). Religion and Spirituality: Why and How to Address It in Clinical Practice. *Child and adolescent psychiatric clinics of North America*, 31(4), 615–630. <https://doi.org/10.1016/j.chc.2022.05.007>
- Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2018*. Atlanta: U.S. Department of Health and Human Services; 2019. DOI: 10.15620/cdc.79370.
- Centers for Disease Control and Prevention. (2021). 2023 Sexually Transmitted Disease Surveillance. Retrieved from: <https://www.cdc.gov/std/statistics/2021/default.htm>
- Centers for Disease Control and Prevention. (2023a). *Detailed STD Facts - Chlamydia*. <https://www.cdc.gov/std/chlamydia/stdfact-chlamydia-detailed.htm>
- Centers for Disease Control and Prevention. (2023b). *Detailed STD facts - Gonorrhea*. Centers for Disease Control and Prevention. <https://www.cdc.gov/std/gonorrhea/stdfact-gonorrhea-detailed.htm#:~:text=What%20is%20gonorrhea%3F,urethra%20in%20women%20and%20men.>
- Centers for Disease Control and Prevention. (2023c). *Detailed STD facts - Syphilis*. Centers for Disease Control and Prevention. <https://www.cdc.gov/std/syphilis/stdfact-syphilis-detailed.htm>
- Centers for Disease Control and Prevention. (n.d.) *Surveillance Case Definitions for Current and Historical Conditions*. Retrieved from: <https://ndc.services.cdc.gov>.
- Charendoff, R. (2019). Navigating the Laws of Shomer Negiah in a Secular World. <https://doi.org/10.21985/N2XJ4X>

- Culotta, E. (2009). On the Origin of Religion. *Science*, 326(5954), 784–787.
- Dow, J. W. (2006). The Evolution of Religion: Three Anthropological Approaches. *Method & Theory in the Study of Religion*, 18(1), 67–91.
- Durkheim, Emile. (2005[1897]). Suicide, a study in sociology. J. A. Spaulding, & G. Simpson (trans.). London: Routledge. <https://www.gacbe.ac.in/images/E%20books/Durkheim%20-%20Suicide%20-%20A%20study%20in%20sociology.pdf>
- Durkheim, Emile. (2012[1915]). The Elementary Forms of the Religious Life. Joseph Ward Swain (trans.). Project Gutenberg. <https://www.gutenberg.org/files/41360/41360-h/41360-h.htm>
- Farley, T. A. (2006). Sexually Transmitted Diseases in the Southeastern United States: Location, Race, and Social Context. *Sexually Transmitted Diseases*, 33(7), S58. <https://doi.org/10.1097/01.olq.0000175378.20009.5a>
- Flemming R. (The Wrong Kind of) Gonorrhoea in Antiquity. In: Szreter S, editor. The Hidden Affliction: Sexually Transmitted Infections and Infertility in History. Rochester (NY): University of Rochester Press; 2019 Oct. Chapter One. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK547155/>
- Fu, L., Sun, Y., Han, M., Wang, B., Xiao, F., Zhou, Y., Gao, Y., Fitzpatrick, T., Yuan, T., Li, P., Zhan, Y., Lu, Y., Luo, G., Duan, J., Hong, Z., Fairley, C. K., Zhang, T., Zhao, J., & Zou, H. (2022). Incidence Trends of Five Common Sexually Transmitted Infections Excluding HIV From 1990 to 2019 at the Global, Regional, and National Levels: Results From the Global Burden of Disease Study 2019. *Frontiers in medicine*, 9, 851635. <https://doi.org/10.3389/fmed.2022.851635>
- Gelfand, T. The History of the Medical Profession. In: Bynum, W. F., & Porter, R. (Eds.). (1993). *Companion encyclopedia of the history of medicine*. Taylor & Francis Group.
- Ghossoub, E., Kassir, G., El Bashour, J., & Saneh, W. (2022). Associations between religiosity, aggression and crime: Results from the National Survey on Drug Use and Health. *Social Psychiatry and Psychiatric Epidemiology*, 57(9), 1829–1838. <https://doi.org/10.1007/s00127-021-02181-y>
- Grafton, T. H. (1945). Religious Origins and Sociological Theory. *American Sociological Review*, 10(6), 726–739. <https://doi.org/10.2307/2085842>
- Gruber, F., Lipozenčić, J., & Kehler, T. (2015). History of venereal diseases from antiquity to the renaissance. *Acta dermatovenerologica Croatica: ADC*, 23(1), 1–11.
- Hall, K. S., Moreau, C., & Trussell, J. (2012). Lower use of sexual and reproductive health services among women with frequent religious participation, regardless of sexual experience. *Journal of women's health (2002)*, 21(7), 739–747. <https://doi.org/10.1089/jwh.2011.3356>

Hamilton, D. T., Katz, D. A., Haderxhanaj, L. T., Copen, C. E., Spicknall, I. H., & Hogben, M. (2023). Modeling the impact of changing sexual behaviors with opposite-sex partners and STI testing among women and men ages 15–44 on STI diagnosis rates in the United States 2012–2019. *Infectious Disease Modelling*, 8(4), 1169–1176. <https://doi.org/10.1016/j.idm.2023.10.005>

Hemmati, R., Bidel, Z., Nazarzadeh, M., Valadi, M., Berenji, S., Erami, E., Al Zaben, F., Koenig, H. G., Sanjari Moghaddam, A., Teymoori, F., Sabour, S., Ghanbarizadeh, S. R., & Seghatoleslam, T. (2019). Religion, Spirituality and Risk of Coronary Heart Disease: A Matched Case-Control Study and Meta-Analysis. *Journal of Religion and Health*, 58(4), 1203–1216. <https://doi.org/10.1007/s10943-018-0722-z>

Henning, C. L. (1898). On the Origin of Religion. *American Anthropologist*, 11(12), 373–382.

Hernigou, P. (2013). Ambroise Paré’s life (1510–1590): Part I. *International Orthopaedics*, 37(3), 543–547. <https://doi.org/10.1007/s00264-013-1797-5>

Institute for Health Metrics and Evaluation (IHME). GBD Compare. Seattle, WA: IHME, University of Washington, 2015. Available from <http://vizhub.healthdata.org/gbd-compare>. (Accessed August 17, 2023)

Koenig, H. G. (2000). Religion and medicine I: Historical background and reasons for separation. *International Journal of Psychiatry in Medicine*, 30(4), 385–398. <https://doi.org/10.2190/2RWB-3AE1-M1E5-TVHK>

Koenig, H. G. (2012). Religion, Spirituality, and Health: The Research and Clinical Implications. *ISRN Psychiatry*, 2012, 278730. <https://doi.org/10.5402/2012/278730>

Kreisel, K. M., Spicknall, I. H., Gargano, J. W., Lewis, F. M., Lewis, R. M., Markowitz, L. E., Roberts, H., Johnson, A. S., Song, R., Cyr, S. B. St., Weston, E. J., Torrone, E. A., & Weinstock, H. S. (2021). Sexually Transmitted Infections Among US Women and Men: Prevalence and Incidence Estimates, 2018. *Sexually Transmitted Diseases*, 48(4), 208–214. <https://doi.org/10.1097/OLQ.0000000000001355>

Kristeller, P. O. (1978). Humanism. *Minerva*, 16(4), 586–595.

Leichliter JS, Seiler N, Wohlfeiler D. Sexually Transmitted Disease Prevention Policies in the United States: Evidence and Opportunities. *Sexually Transmitted Diseases*. 2016;43(2 Suppl 1): S113-S121. doi:10.1097/OLQ.0000000000000289

Levin, J. (2022). Toward a translational epidemiology of religion: Challenges and applications. *Annals of Epidemiology*, 75, 25–31. <https://doi.org/10.1016/j.annepidem.2022.08.053>

Lieberman, J. A., Cannon, C. A., & Bourassa, L. A. (2021). Laboratory Perspective on Racial Disparities in Sexually Transmitted Infections. *The journal of applied laboratory medicine*, 6(1), 264–273. <https://doi.org/10.1093/jalm/jfaa163>

Lipka, M., & Gecewicz, C. (2017). More Americans now say they're spiritual but not religious. *Pew Research Center*. Retrieved August 17, 2023, from <https://www.pewresearch.org/short-reads/2017/09/06/more-americans-now-say-theyre-spiritual-but-not-religious/>

Longhofer, W., & Winchester, D. (Eds.). (2016). *Social theory re-wired: New connections to classical and contemporary perspectives*. Taylor & Francis Group.

Marfatia, Y. S., Pandya, I., & Mehta, K. (2015). Condoms: Past, present, and future. *Indian Journal of Sexually Transmitted Diseases and AIDS*, *36*(2), 133–139. <https://doi.org/10.4103/2589-0557.167135>

Miller, W. R., & Thoresen, C. E. (2003). Spirituality, religion, and health. An emerging research field. *The American Psychologist*, *58*(1), 24–35. <https://doi.org/10.1037/0003-066x.58.1.24>

Mojola, S. A., & Everett, B. (2012). STD and HIV risk factors among U.S. young adults: variations by gender, race, ethnicity and sexual orientation. *Perspectives on sexual and reproductive health*, *44*(2), 125–133. <https://doi.org/10.1363/4412512>

Moon, J. W., Cohen, A. B., Laurin, K., & MacKinnon, D. P. (2023). Is Religion Special? *Perspectives on Psychological Science*, *18*(2), 340–357. <https://doi.org/10.1177/17456916221100485>

Moons, P., Luyckx, K., Dezutter, J., Kovacs, A. H., Thomet, C., Budts, W., Enomoto, J., Sluman, M. A., Yang, H.-L., Jackson, J. L., Khairy, P., Subramanyan, R., Alday, L., Eriksen, K., Dellborg, M., Berghammer, M., Johansson, B., Mackie, A. S., Menahem, S., ... International Society for Adult Congenital Heart Disease (ISACHD). (2019). Religion and spirituality as predictors of patient-reported outcomes in adults with congenital heart disease around the globe. *International Journal of Cardiology*, *274*, 93–99. <https://doi.org/10.1016/j.ijcard.2018.07.103>

Nair, D., Cavanaugh, K. L., Wallston, K. A., Mason, O., Stewart, T. G., Blot, W. J., Ikizler, T. A., & Lipworth, L. P. (2020). Religion, Spirituality, and Risk of End-Stage Kidney Disease Among Adults of Low Socioeconomic Status in the Southeastern United States. *Journal of Health Care for the Poor and Underserved*, *31*(4), 1727–1746. <https://doi.org/10.1353/hpu.2020.0129>

National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Population Health and Public Health Practice; Committee on Prevention and Control of Sexually Transmitted Infections in the United States, Crowley, J. S., Geller, A. B., & Vermund, S. H. (Eds.). (2021). *Sexually Transmitted Infections: Adopting a Sexual Health Paradigm*. National Academies Press (US).

Nowotny, K. M., Omori, M., McKenna, M., & Kleinman, J. (2020). Incarceration Rates and Incidence of Sexually Transmitted Infections in US Counties, 2011–2016. *American Journal of Public Health*, *110*(Suppl 1), S130–S136. <https://doi.org/10.2105/AJPH.2019.305425>



Nugraheni, S. E., & Hastings, J. F. (2020). Relationship between Religious Support and Major Depressive Episode for Adult Non-Medical Prescription Opioid Users and Non-Users. *Substance Use & Misuse*, 55(4), 564–571. <https://doi.org/10.1080/10826084.2019.1688351>

Owusu-Edusei, K., McClendon-Weary, B., Bull, L., Gift, T. L., & Aral, S. O. (2020). County-Level Social Capital and Bacterial Sexually Transmitted Infections in the United States. *Sexually Transmitted Diseases*, 47(3), 165–170. <https://doi.org/10.1097/OLQ.0000000000001117>

Parker, MA., Anthony, JC. (2015). Epidemiological evidence on extra-medical use of prescription pain relievers: transitions from newly incident use to dependence among 12-21 year olds in the United States using meta-analysis, 2002-2011. *PeerJ*:e1340. <https://doi.org/10/7717/peerj.1340>

Peeling RW, Mabey D, Kamb ML, Chen XS, Radolf JD, Benzaken AS. Syphilis. *Nature Reviews Disease Primers*. 2017;3:17073. Published 2017 Oct 12. doi:10.1038/nrdp.2017.73

Pew Research Center, September, 2022, “Modeling the Future of Religion in America”. Retrieved from: <https://www.pewresearch.org/religion/2022/09/13/how-u-s-religious-composition-has-changed-in-recent-decades/>

Pollack, C. C., Bradburne, J., Lee, N. K., Manabe, Y. C., Widdice, L. E., Gaydos, C. A., Tuddenham, S. A., Rompalo, A. M., Jackman, J., & Timm, C. M. (2023). A National, County-Level Evaluation of the Association Between COVID-19 and Sexually Transmitted Infections Within the United States in 2020. *Sexually Transmitted Diseases*, 50(8), 536–542. <https://doi.org/10.1097/OLQ.0000000000001818>

Risse, G.B. (1993). Medical Care. In Bynum, W. F., & Porter, R. (Eds.). (1993). Companion encyclopedia of the history of medicine. Taylor & Francis Group.

Royston P, Sauerbrei W 2008: Multivariable model-building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, Chichester: John Wiley & Sons Ltd. ISBN 978-0-470-02842-1

Rudowitz, R., Drake, P., Tolbert, J., & Damico, A. (2023). *How many uninsured are in the coverage gap and how many could be eligible if all states adopted the Medicaid expansion?*. KFF. <https://www.kff.org/medicaid/issue-brief/how-many-uninsured-are-in-the-coverage-gap-and-how-many-could-be-eligible-if-all-states-adopted-the-medicare-expansion/>

Saroglou, V. (2011). Believing, Bonding, Behaving, and Belonging: The Big Four Religious Dimensions and Cultural Variation. *Journal of Cross-Cultural Psychology*, 42(8), 1320–1340. <https://doi.org/10.1177/0022022111412267>

Salas-Wright, C. P., Vaughn, M. G., Ugalde, J., & Todic, J. (2015). Substance use and teen pregnancy in the United States: Evidence from the NSDUH 2002-2012. *Addictive Behaviors*, 45, 218–225. <https://doi.org/10.1016/j.addbeh.2015.01.039>

Sexuality Information and Education Council of the United States. 2022. Sex Ed State Law and Policy Chart SIECUS State Profiles: July 2022. Retrieved from: <https://siecus.org/wp-content/uploads/2021/09/2022-Sex-Ed-State-Law-and-Policy-Chart.pdf>

Sisti, L. G., Buonsenso, D., Moscato, U., Costanzo, G., & Malorni, W. (2023). The Role of Religions in the COVID-19 Pandemic: A Narrative Review. *International Journal of Environmental Research and Public Health*, 20(3), 1691. <https://doi.org/10.3390/ijerph20031691>

Stafford IA, Berra A, Minard CG, et al. Challenges in the Contemporary Management of Syphilis among Pregnant Women in New Orleans, LA. *Infectious Diseases in Obstetrics and Gynecology*. 2019;2019:2613962. Published 2019 Feb 13. doi:10.1155/2019/2613962

StataCorp. 2023. *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC.

Substance Abuse and Mental Health Services Administration. Reliability of Key Measures in the National Survey on Drug Use and Health [Internet]. Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2010 Feb. Appendix A, Description of NSDUH. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK519790/>

Tampa, M., Sarbu, I., Matei, C., Benea, V., & Georgescu, S. R. (2014). Brief history of syphilis. *Journal of medicine and life*, 7(1), 4–10.

Tansey, E.M. The Physiological Tradition. In: Bynum, W. F., & Porter, R. (Eds.). (1993). *Companion encyclopedia of the history of medicine*. Taylor & Francis Group.

Unemo, M., Bradshaw, C. S., Hocking, J. S., de Vries, H. J. C., Francis, S. C., Mabey, D., Marrazzo, J. M., Sonder, G. J. B., Schwebke, J. R., Hoornenborg, E., Peeling, R. W., Philip, S. S., Low, N., & Fairley, C. K. (2017). Sexually transmitted infections: challenges ahead. *The Lancet. Infectious diseases*, 17(8), e235–e279. [https://doi.org/10.1016/S1473-3099\(17\)30310-9](https://doi.org/10.1016/S1473-3099(17)30310-9)

US Census Bureau. (2000). 2000 Census Block Maps. Retrieved from: <https://www.census.gov/geographies/reference-maps/2000/geo/2000-census-block-maps.html>

Watts, F. (2020). The evolution of religious cognition. *Archive for the Psychology of Religion*, 42(1), 89–100. <https://doi.org/10.1177/0084672420909479>

Weber, Samuel R., Pargament, Kenneth I. The role of religion and spirituality in mental health. *Current Opinion in Psychiatry* 27(5): p 358-363, September 2014. DOI: 10.1097/YCO.0000000000000080

Westreich, D., & Greenland, S. (2013). The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology*, 177(4), 292–298. <https://doi.org/10.1093/aje/kws412>

Worboys, M. (2019). Chlamydia: A Disease without a History. In S. Szreter (Ed.), *The Hidden Affliction: Sexually Transmitted Infections and Infertility in History*. University of Rochester Press.

Wunn, I. (2000). Beginning of Religion. *Numen*, 47(4), 417–452.

Zheng, Y., Yu, Q., Lin, Y., Zhou, Y., Lan, L., Yang, S., & Wu, J. (2022). Global burden and trends of sexually transmitted infections from 1990 to 2019: an observational trend study. *The Lancet. Infectious diseases*, 22(4), 541–551. [https://doi.org/10.1016/S1473-3099\(21\)00448-5](https://doi.org/10.1016/S1473-3099(21)00448-5)

## APPENDIX A

**Table A1.** Diagnostic criteria from CDC for each of the three nationally notifiable sexually transmitted diseases. Cases of each of the three diseases must meet both the clinical description as well as the laboratory criteria to be confirmed for reporting purposes.

Chlamydia	
<p><b>Clinical Description:</b> Infection with <i>Chlamydia trachomatis</i> may result in urethritis, epididymitis, cervicitis, acute salpingitis, or other syndromes when sexually transmitted; however, the infection is often asymptomatic in women. Perinatal infections may result in inclusion conjunctivitis and pneumonia in newborns. Other syndromes caused by <i>C. trachomatis</i> include lymphogranuloma venereum (see Lymphogranuloma venereum) and trachoma.</p>	<p><b>Laboratory Criteria:</b> Isolation of <i>C. trachomatis</i> by culture, <b>OR</b> Demonstration of <i>C. trachomatis</i> in a clinical specimen by detection of antigen or nucleic acid</p>
Gonorrhea	
<p><b>Clinical Description:</b> A sexually transmitted infection commonly manifested by urethritis, cervicitis, proctitis, salpingitis, or pharyngitis. Infection may be asymptomatic.</p>	<p><b>Laboratory Criteria:</b> Observation of gram-negative intracellular diplococci in a urethral smear obtained from a male or an endocervical smear obtained from a female, <b>OR</b> isolation of typical gram-negative, oxidase-positive diplococci by culture (presumptive <i>Neisseria gonorrhoeae</i>) from a clinical specimen, <b>OR</b> demonstration of <i>N. gonorrhoeae</i> in a clinical specimen by detection of antigen or nucleic acid</p>
Syphilis	
Primary	
<p><b>Clinical Description:</b> A stage of infection with <i>Treponema pallidum</i> characterized by one or more ulcerative lesions (e.g. chancre), which might differ considerably in clinical appearance.</p>	<p><b>Laboratory Criteria:</b> Demonstration of <i>T. pallidum</i> by darkfield microscopy in a clinical specimen that was not obtained from the oropharynx and is not potentially contaminated by stool, <b>OR</b> Demonstration of <i>T. pallidum</i> by polymerase chain reaction (PCR) or equivalent direct molecular methods in any clinical specimen</p>
Secondary	
<p><b>Clinical Description:</b> A stage of infection caused by <i>T. pallidum</i> characterized by localized or diffuse mucocutaneous lesions (e.g., rash – such as non-pruritic macular,</p>	<p><b>Laboratory Criteria:</b> Demonstration of <i>T. pallidum</i> by darkfield microscopy in a clinical specimen that was not obtained from the oropharynx and is not potentially</p>

**Table A1 (cont'd.)**

maculopapular, papular, or pustular lesions), often with generalized lymphadenopathy. Other signs can include mucous patches, condyloma lata, and alopecia. The primary ulcerative lesion may still be present.	contaminated by stool, <b>OR</b> Demonstration of <i>T. pallidum</i> by polymerase chain reaction (PCR) or equivalent direct molecular methods in any clinical specimen
Early non-primary non-secondary	
<b>Clinical Description:</b> A stage of infection caused by <i>T. pallidum</i> in which initial infection has occurred within the previous 12 months, but there are no signs or symptoms of primary or secondary syphilis	<b>Laboratory Criteria:</b> A current nontreponemal test titer demonstrating fourfold or greater increase from the last nontreponemal test titer, unless there is evidence that this increase was not sustained for >2 weeks.

**Table A2.** Comparison of generalized linear regression of log transformed total STD cases and negative binomial regression of total STD cases. Significance of covariates remains the same regardless of model except for in the case of syphilis.

	Coefficient	Std. err.	z	P>z	[95% conf. interval]
Chlamydia					
Religiosity	1.44	0.70	2.06	0.039	0.07, 2.80
logChlamydia					
Religiosity	2.53	0.69	3.67	0.000	1.18, 3.88
Gonorrhea					
Religiosity	2.31	0.81	2.84	0.005	0.72, 3.90
logGonorrhea					
Religiosity	5.08	0.88	5.73	<0.001	3.34, 6.81
Syphilis					
Religiosity	-0.08	0.95	-0.08	0.937	-1.94, 1.79
logSyphilis					
Religiosity	2.94	1.05	2.80	0.005	0.88, 5.00
Total STIs					
Religiosity	1.59	0.72	2.21	0.027	0.18, 3.00
logSTD					
Religiosity	2.92	0.72	4.05	<0.001	1.51, 4.34

**Table A3.** Example negative binomial regression of chlamydia cases by level of religiosity (top) and the zero truncated negative binomial regression of the same variables (bottom). No appreciable difference was observed between the two methods.

Chlamydia	Coefficient	Std. err.	z	P>z	[95% conf. interval]	
Negative binomial						
Religiosity	1.44	.70	2.06	0.039	.07	2.80
Zero truncated negative binomial						
Religiosity	1.44	.70	2.06	0.039	.07	2.80

**Table A4.** mean and variance of the four STD outcomes. The much larger variance for each outcome points to the presence of overdispersion in the data, suggesting the use of a negative binomial model as opposed to a Poisson approach.

Outcome	Mean	Variance
Chlamydia	32676.13	1.35e+09
Gonorrhea	9351.576	1.44e+08
Syphilis	1073.254	3760842
Total STI cases	43100.96	2.52e+09

**Table A5.** Additional covariates included in Aim 3 models with the operational definition of the variable as recoded by the investigator.

Variable Name	Description	Operationalized
Non-Hispanic White	White identity	Proportion of individuals identifying as white in the substate region
Hispanic ethnicity	Are they of Hispanic origin – imputation revised	Proportion of individuals identifying as Hispanic in the substate region
Gender	Gender – imputation revised	Proportion male of the substate region
Alcohol use disorder (AUD)	During the past 12 months, there was alcohol dependence or abuse (DSM-IV) – recoded	Proportion who responded YES in the substate region
Drug use disorder (DUD)	During the past 12 months, there was illicit drug abuse (DSM-IV) – recoded	Proportion who responded YES in the substate region
Poverty	Living below the poverty level	Proportion of the substate region living at or below 100% of the US Census poverty threshold
Education	Educational attainment	Proportion of the substate region with less than a high school education
Employment	Employment status	Proportion employed full time in the substate region

**Table A5 (cont'd.)**

Insurance status	Any health insurance coverage reported	Proportion uninsured in the substate region
Booked for a crime	Ever in their life been arrested and booked for breaking the law	Proportion who answered YES in the substate region
Tobacco usage	During the past 30 days, did they use any tobacco – recoded	Proportion who responded YES in the substate region
Psychotherapeutic use	During the past 12 months, if they used psychotherapeutics	Proportion who responded YES in the substate region

**Table A6.** Effect of individual covariate addition to the population-adjusted crude model on the slope of the religiosity – chlamydia outcome relationship.

Chlamydia model +	Slope (95% CI)	LR test	P > chi <sup>2</sup>
Religiosity	1.44 (0.07, 2.80)	--	--
Religiosity + population	2.12 (1.34, 2.89)	381.50	<0.001
Religiosity + population + age	2.08 (1.31, 2.85)	49.06	<0.001
Religiosity + population + race/ethnicity	1.06 (0.34, 1.78)	104.14	<0.001
Religiosity + population + gender	1.88 (1.11, 2.66)	17.65	<0.001
Religiosity + population + AUD	2.77 (1.90, 3.64)	9.17	0.0025
Religiosity + population + DUD	3.09 (2.33, 3.84)	66.00	<0.001
Religiosity + population + poverty	0.93 (0.07, 1.80)	30.45	<0.001
Religiosity + population + education	1.29 (0.37, 2.20)	10.72	0.0011
Religiosity + population + employment	2.10 (1.32, 2.89)	0.05	0.8273
Religiosity + population + insurance coverage	1.70 (0.83, 2.57)	4.40	0.0359
Religiosity + population + booked for a crime	2.11 (1.34, 2.88)	2.23	0.1352
Religiosity + population + tobacco usage	2.27 (1.45, 3.08)	1.40	0.2370
Religiosity + population + extra-medical drug use	2.47 (1.70, 3.25)	22.78	<0.001

**Table A7.** Effect of individual covariate addition to the population-adjusted crude model on the slope of the religiosity – gonorrhea outcome relationship.

Gonorrhea model +	Slope (95% CI)	LR test	P > chi <sup>2</sup>
Religiosity	2.31 (0.72, 3.90)	--	--
Religiosity + population	3.41 (2.34, 4.49)	267.63	<0.001

**Table A7 (cont'd.)**

Religiosity + population + age	4.11 (3.01, 5.21)	60.78	<0.001
Religiosity + population + race/ethnicity	2.18 (1.17, 3.20)	116.38	<0.001
Religiosity + population + gender	3.17 (2.11, 4.24)	19.42	<0.001
Religiosity + population + AUD	5.31 (4.07, 6.54)	27.64	<0.001
Religiosity + population + DUD	5.45 (4.39, 6.52)	92.50	<0.001
Religiosity + population + poverty	1.95 (0.73, 3.18)	22.41	<0.001
Religiosity + population + education	2.80 (1.48, 4.13)	2.38	0.1229
Religiosity + population + employment	3.53 (2.44, 4.63)	1.29	0.2564
Religiosity + population + insurance coverage	3.08 (1.85, 4.31)	1.16	0.2807
Religiosity + population + booked for a crime	3.33 (2.28, 4.39)	4.36	0.0367
Religiosity + population + tobacco usage	3.31 (2.16, 4.46)	0.23	0.6344
Religiosity + population + extra-medical drug use	4.42 (3.32, 5.51)	40.76	<0.001

**Table A8.** Effect of individual covariate addition to the population-adjusted crude model on the slope of the religiosity – syphilis outcome relationship.

Syphilis model +	Slope (95% CI)	LR test	P > chi <sup>2</sup>
Religiosity	-0.08 (-1.94, 1.79)	--	--
Religiosity + population	1.02 (-0.12, 2.16)	287.15	<0.001
Religiosity + population + age	2.20 (1.01, 3.39)	93.50	<0.001
Religiosity + population + race/ethnicity	-0.49 (-1.49, 0.51)	204.04	<0.001
Religiosity + population + gender	0.89 (-0.25, 2.02)	25.34	<0.001
Religiosity + population + AUD	3.35 (1.97, 4.74)	27.00	<0.001
Religiosity + population + DUD	3.94 (2.79, 5.08)	110.16	<0.001
Religiosity + population + poverty	-1.16 (-2.46, 0.13)	40.67	<0.001
Religiosity + population + education	-0.45 (-1.83, 0.92)	12.66	0.0004



**Table A8 (cont'd.)**

Religiosity + population + employment	1.34 (0.16, 2.53)	4.78	0.0288
Religiosity + population + insurance coverage	-0.58 (-1.90, 0.73)	20.11	<0.001
Religiosity + population + booked for a crime	1.03 (-0.12, 2.18)	0.06	0.8034
Religiosity + population + tobacco usage	1.94 (0.70, 3.18)	13.87	0.0002
Religiosity + population + extra-medical drug use	2.51 (1.30, 3.73)	38.42	<0.001

**Table A9.** Result of the plausible predictor model for total substate region STI cases.

Total STIs	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	1.86	0.55	3.37	0.001	0.78, 2.93
Population	0.81	0.05	17.36	<0.001	0.72, 0.91
% age 12-17	-0.68	2.40	-0.28	0.777	-5.39, 4.03
% age 18-25	1.64	1.49	1.10	0.270	-1.28, 4.56
% age 26-35	3.93	1.42	2.76	0.006	1.14, 6.72
% age 36-49	-0.58	1.42	-0.41	0.683	-3.36, 2.20
Non-Hispanic White	-1.92	0.30	-6.39	<0.001	-2.51, -1.33
Hispanic ethnicity	-1.45	0.3749125	-3.88	<0.001	-2.19, -0.72
Gender	-3.26	1.38	-2.36	0.018	-5.96, -0.55
AUD	0.91	2.09	0.43	0.664	-3.19, 5.01
DUD	16.03	4.19	3.83	<0.001	7.82, 24.25
Poverty	-0.45	1.13	-0.39	0.693	-2.67, 1.77
Education	0.71	1.18	0.60	0.551	-1.62, 3.03

**Table A10.** Results of the plausible predictors model for chlamydia cases.

Chlamydia	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	1.42	0.52	2.73	0.006	0.40, 2.44
Population	0.79	0.04	18.01	<0.001	0.71, 0.88
% age 12-17	-0.10	2.29	-0.04	0.965	-4.59, 4.40
% age 18-25	1.89	1.41	1.34	0.179	-0.87, 4.65
% age 26-35	3.49	1.36	2.57	0.010	0.83, 6.15
% age 36-49	-0.41	1.34	-0.31	0.759	-3.04, 2.22
Non-Hispanic White	-1.70	0.28	-6.09	<0.001	-2.25, -1.15
Hispanic ethnicity	-1.24	0.35	-3.51	<0.001	-1.93, -0.55
Gender	-3.08	1.31	-2.35	0.019	-5.64, -0.52
AUD	0.44	1.97	0.22	0.823	-3.43, 4.31

**Table A10 (cont'd.)**

DUD	13.54	3.98	3.40	0.001	5.73, 21.35
Poverty	-0.38	1.07	-0.35	0.724	-2.48, 1.73
Education	0.84	1.12	0.75	0.451	-1.35, 3.04

**Table A11.** Results of the plausible predictors model for gonorrhea cases.

Gonorrhea	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	4.04	0.75	5.40	<0.001	2.57, 5.50
Population	0.89	0.06	14.09	<0.001	0.76, 1.01
% age 12-17	-1.00	3.14	-0.32	0.750	-7.16, 5.15
% age 18-25	0.74	1.97	0.37	0.708	-3.13, 4.61
% age 26-35	4.88	1.84	2.66	0.008	1.29, 8.48
% age 36-49	-1.78	1.88	-0.94	0.345	-5.46, 1.91
Non-Hispanic White	-2.65	0.42	-6.32	<0.001	-3.48, -1.83
Hispanic ethnicity	-2.28	0.51	-4.46	<0.001	-3.27, -1.28
Gender	-3.85	1.82	-2.12	0.034	-7.41, -0.28
AUD	3.05	2.82	1.08	0.278	-2.46, 8.57
DUD	25.16	5.45	4.61	<0.001	14.47, 35.85
Poverty	-0.98	1.50	-0.65	0.516	-3.92, 1.97
Education	-0.08	1.59	-0.05	0.962	-3.19, 3.04

**Table A12.** Results of the plausible predictors model for primary, secondary, and early non-primary non-secondary syphilis cases.

Syphilis	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	1.25	0.75	1.66	0.096	-0.22, 2.72
Population	0.94	0.07	13.89	<0.001	0.81, 1.07
% age 12-17	-9.927	3.32	-2.99	0.003	-16.43, -3.42
% age 18-25	0.67	2.02	0.33	0.742	-3.30, 4.64
% age 26-35	6.94	1.93	3.60	<0.001	3.16, 10.72
% age 36-49	-1.64	1.94	-0.84	0.398	-5.44, 2.16
Non-Hispanic White	-4.04	0.44	-9.09	<0.001	-4.92, -3.17
Hispanic ethnicity	-1.69	0.56	-3.01	0.003	-2.79, -0.59
Gender	-3.46	1.91	-1.81	0.071	-7.21, -0.29
AUD	-3.24	2.89	-1.12	0.262	-8.90, 2.42
DUD	27.89	5.82	4.79	<0.001	16.46, 39.29
Poverty	-2.12	1.55	-1.37	0.171	-5.17, 0.92
Education	0.76	1.64	0.46	0.644	-2.45, 3.96

**Table A13.** Results of the unlikely predictor model for total STI cases in the US

Total STIs	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
------------	-------------	-----------	---	-----	----------------------

**Table A13 (cont'd.)**

Religiosity	1.89	0.50	3.78	<0.001	0.91, 2.86
Population	0.78	0.04	17.71	<0.001	0.69, 0.86
% age 26-35	5.08	1.52	3.35	0.001	2.11, 8.05
Non-Hispanic White	-2.39	0.30	-7.95	<0.001	-2.98, -1.80
Hispanic ethnicity	-1.02	0.39	-2.65	0.008	-1.78, -0.27
Gender	-2.93	1.36	-2.16	0.031	-5.58, -0.27
DUD	3.86	4.89	0.79	0.430	-5.72, 13.44
Employment	-1.49	0.74	-2.00	0.045	-2.95, -0.03
Insurance status	-2.49	0.92	-2.69	0.007	-4.30, -0.68
Booked for a crime	-0.20	1.13	-0.18	0.856	-2.41, 2.00
Tobacco usage	1.34	0.80	1.67	0.095	-0.23, 2.92
Extra-medical drug use	8.62	2.29	3.76	<0.001	4.13, 13.12

**Table A14. Results of the unlikely predictor model for chlamydia cases in the US**

Chlamydia	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	1.59	0.47	3.37	0.001	0.66, 2.52
Population	0.76	0.04	18.27	<0.001	0.68, 0.84
% age 26-35	4.70	1.45	3.23	0.001	1.85, 7.54
Non-Hispanic White	-2.13	0.28	-7.58	<0.001	-2.69, -1.58
Hispanic ethnicity	-0.83	0.37	-2.26	0.024	-1.54, -0.11
Gender	-2.78	1.29	-2.16	0.031	-5.30, 0.25
DUD	2.78	4.66	0.60	0.550	-6.34, 11.91
Employment	-1.53	0.71	-2.15	0.031	-2.92, -0.14
Insurance status	-2.26	0.88	-2.57	0.010	-3.98, -0.54
Booked for a crime	-0.11	1.08	-0.11	0.916	-2.22, 1.99
Tobacco usage	1.05	0.76	1.39	0.165	-0.43, 2.54
Extra-medical drug use	7.69	2.19	3.52	<0.001	3.40, 11.97

**Table A15. Results of the unlikely predictor model for gonorrhea cases in the US**

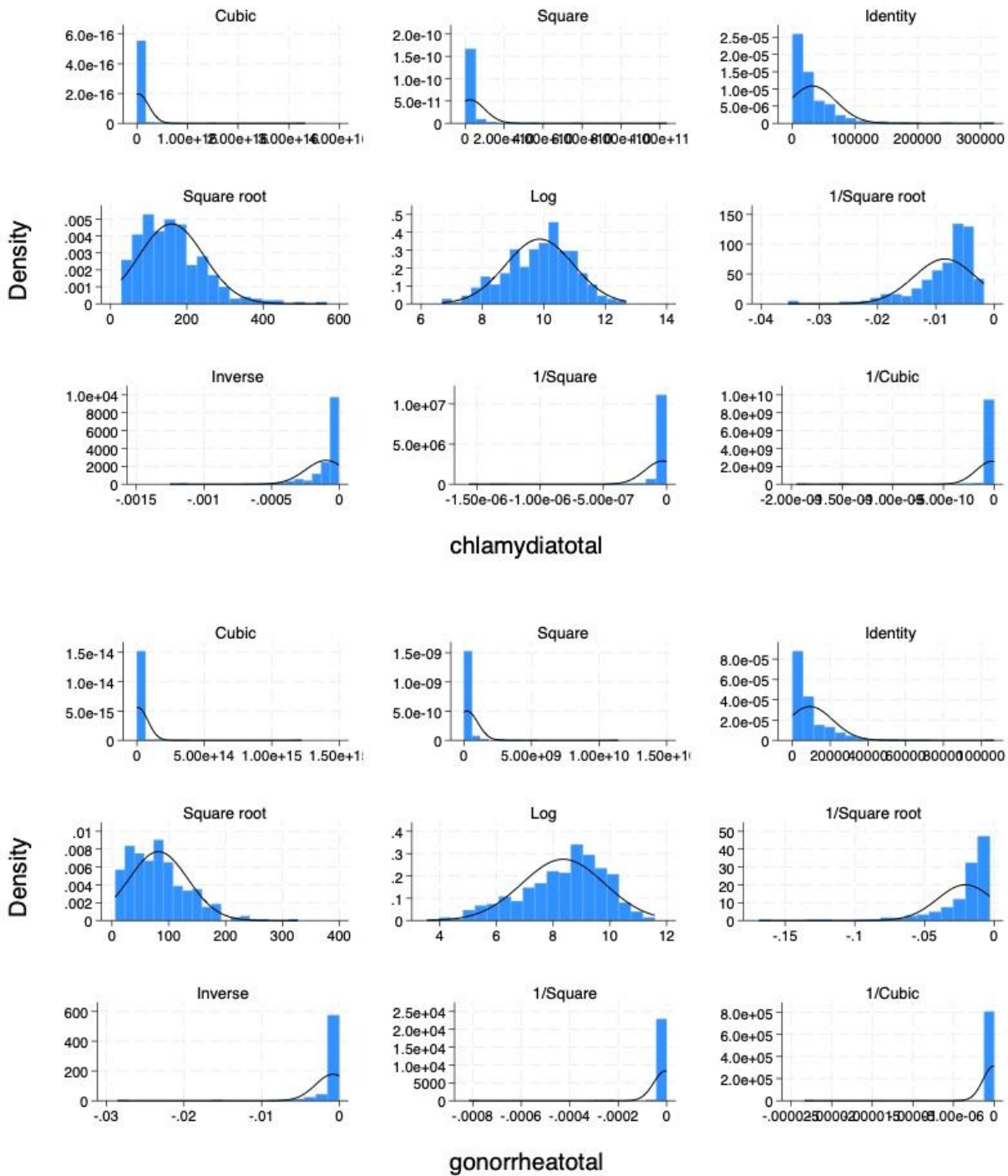
Gonorrhea	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Religiosity	3.58	0.68	5.27	<0.001	2.25, 4.92
Population	0.86	0.06	14.52	<0.001	0.74, 0.97
% age 26-35	6.17	1.96	3.16	0.002*	2.34, 10.01
Non-Hispanic White	-3.30	0.42	-7.95	<0.001	-4.11, -2.49
Hispanic ethnicity	-1.58	0.52	-3.04	0.002*	-2.60, -0.56

**Table A15 (cont'd.)**

Gender	-3.28	1.79	-1.83	0.067	-6.78, 0.22
DUD	6.51	6.35	1.03	0.305	-5.93, 18.95
Employment	-1.56	0.97	-1.61	0.107	-3.45, 0.34
Insurance status	-4.25	1.22	-3.50	<0.001	-6.64, -1.87
Booked for a crime	-0.08	1.47	-0.06	0.955	-2.96, 2.79
Tobacco usage	2.71	1.08	2.51	0.012*	0.60, 4.83
Extra-medical drug use	12.40	2.98	4.16	<0.001	6.56, 18.25

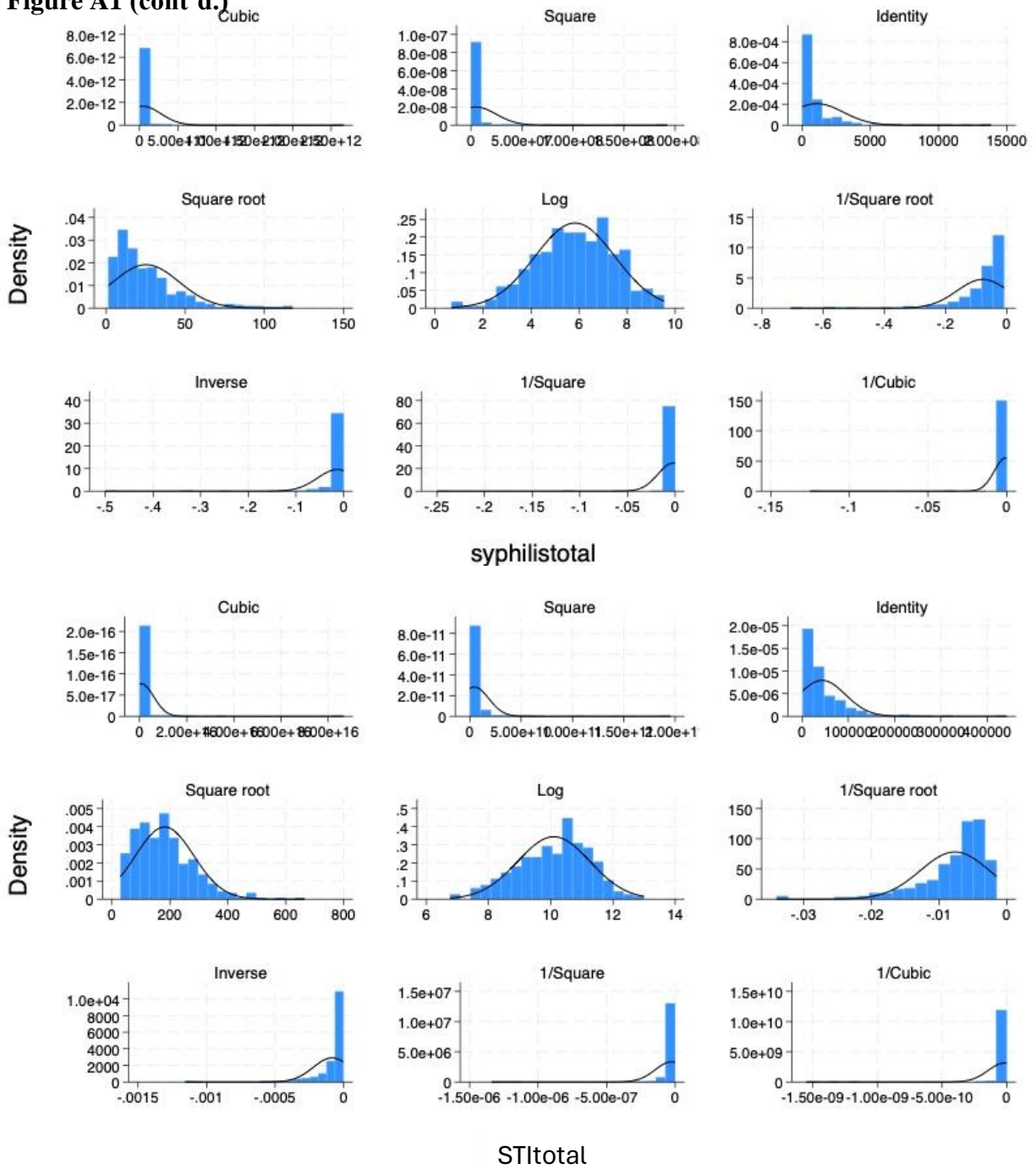
**Table A16.** Results of the unlikely predictor model for primary, secondary, and early non-primary non-secondary syphilis cases in the US

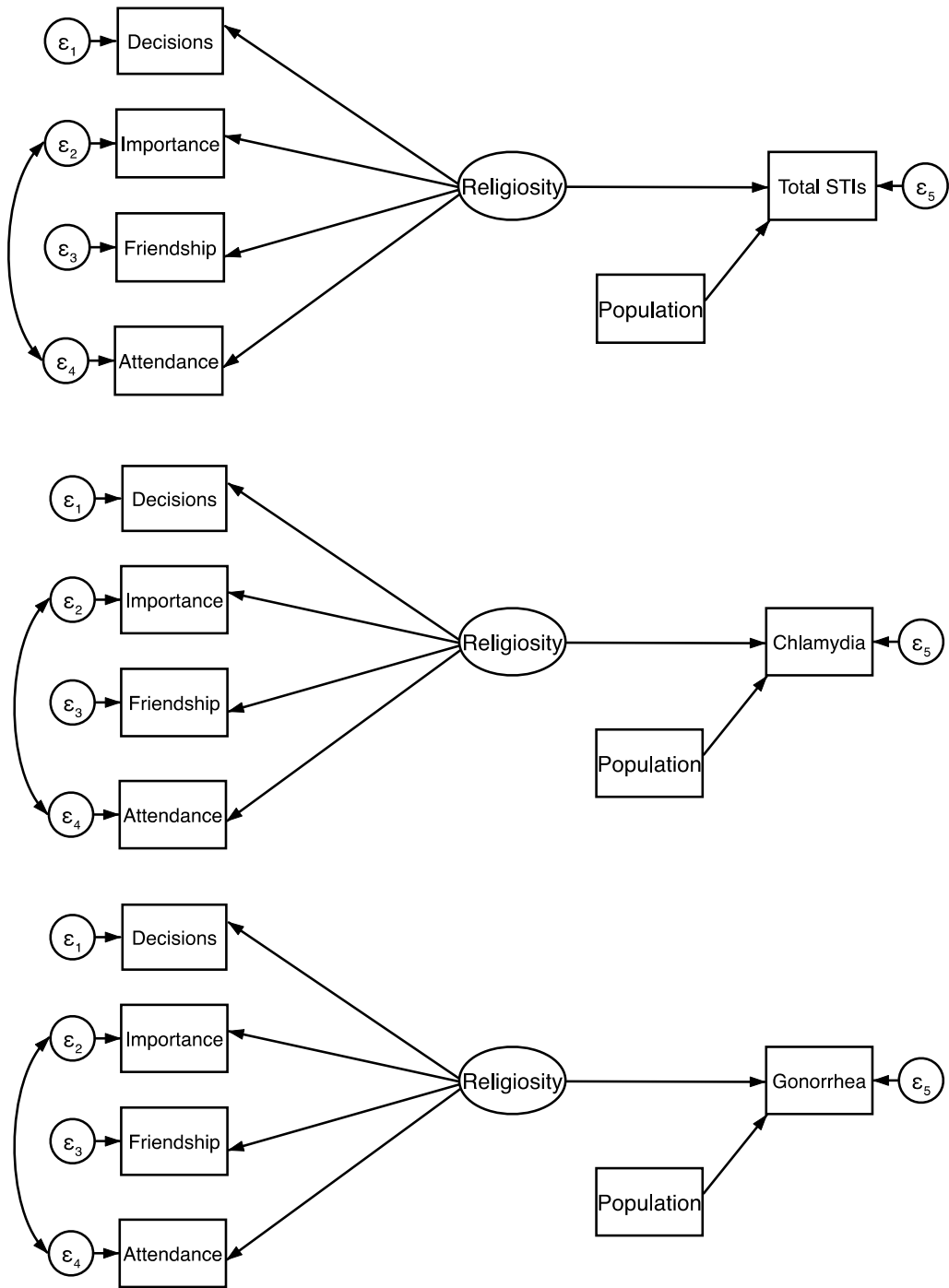
Syphilis	Coefficient	Std. err.	Z	P>z	[95% conf. interval]
Population	8.88	0.06	13.88	<0.001	0.75, 1.00
% age 12-17	-8.85	2.99	-2.96	0.003*	-14.71, -3.00
% age 26-35	6.67	2.07	3.22	0.001*	2.60, 10.73
Non-Hispanic White	-4.72	0.40	-11.87	<0.001	-5.50, -3.94
Hispanic ethnicity	-2.50	0.51	-4.92	<0.001	-3.49, -1.50
DUD	4.98	6.35	0.78	0.434	-7.48, 17.42
Employment	-1.35	1.01	-1.34	0.179	-3.32, 0.62
Insurance status	0.64	1.23	0.53	0.599	-1.76, 3.05
Booked for a crime	-4.84	1.50	-3.21	0.001*	-7.78, -1.89
Tobacco usage	0.41	1.06	0.39	0.697	-1.66, 2.48
Extra-medical drug use	15.06	3.18	4.74	<0.001	8.83, 21.29



**Figure A1.** Histograms showing the fit of all four outcome variables for each of nine different modeling approaches. In all four cases the square root and log transformations appear to provide the best fit to the data. The square root transformation corresponds to the negative binomial model used for analysis while the log transformation corresponds to a general linear model approach.

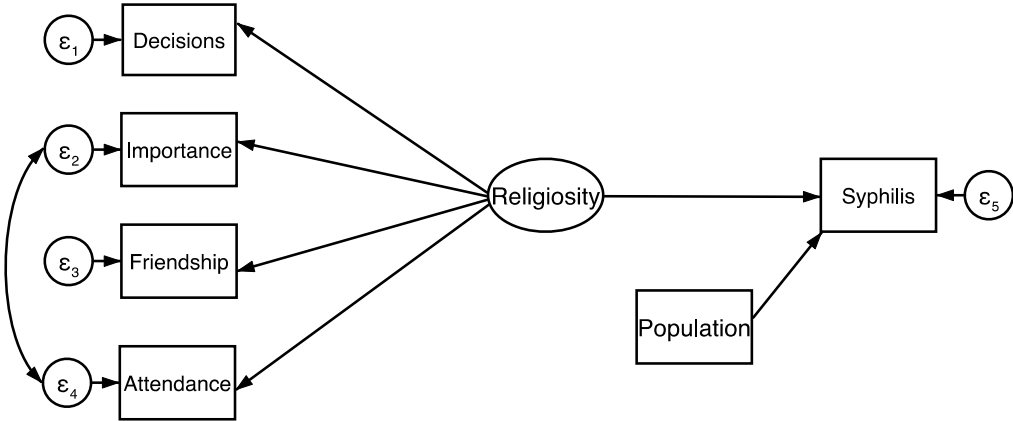
Figure A1 (cont'd.)





**Figure A2.** Cartoon representation of substate region population-adjusted models for all four STI outcomes.

Figure A2 (cont'd.)





## APPENDIX B

```
/* Aim 1 */
```

```
* Get NSDUH religion variables *
```

```
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.01.24/RLGDCSN x STREG10 clean.xlsx", sheet("Sheet 1 - RLGDCSN x STREG10")  
cellrange(A1:M769) firstrow clear
```

```
drop in 1
```

```
drop in 384
```

```
* Removes 'overall' values
```

```
destring J, generate(STATENUM)
```

```
drop I J
```

```
rename L SSRNUM
```

```
generate ssr_n = ((STATENUM*1000)+ SSRNUM)
```

```
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.01.24/Decision.dta"
```

```
* Religion and Friendship
```

```
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.01.24/SNRLFRND x STREG10 clean.xlsx", sheet("Sheet 1 - SNRLFRND x  
STREG10") cellrange(A1:M769) firstrow clear
```

```
drop in 1
```

```
drop in 384
```

```
* Removes 'overall'
```

```
destring J, generate(STATENUM)
```

```

drop I J

rename L SSRNUM

generate ssr_n = ((STATENUM*1000)+ SSRNUM)

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Friendship.dta"

* Religion and Attendance

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/SNRLGSVC x STREG10 clean.xlsx", sheet("Sheet 1 - SNRLGSVC x STREG10-
2") cellrange(A1:M769) firstrow clear

drop in 1

drop in 384

* Removes 'overall'

destring J, generate(STATENUM)

drop I J

rename L SSRNUM

generate ssr_n = ((STATENUM*1000)+ SSRNUM)

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Attendance.dta"

* Religion and Importance

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/SNRLGIMP x STREG10 clean.xlsx", sheet("Sheet 1 - SNRLGIMP x STREG10")
cellrange(A1:M769) firstrow clear

drop in 1

```

```

drop in 384

* Removes 'overall'

destring J, generate(STATENUM)

drop I J

rename L SSRNUM

generate ssr_n = ((STATENUM*1000)+ SSRNUM)

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Importance.dta"

* Separate agree and disagree

* Decision

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Decision.dta"

keep in 1/383

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Decisionagree.dta"

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Decision.dta", clear

keep in 384/766

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Decisiondisagree.dta"

* Friendship

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Friendship.dta", clear

```

keep in 1/383

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Friendshipagree.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Friendship.dta", clear

keep in 384/766

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Friendshipdisagree.dta"

\* Attendance

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Attendance.dta", clear

keep in 1/383

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Attendanceagree.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Attendance.dta", clear

keep in 384/766

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Attendancedisagree.dta"

\*Importance

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.01.24/Importance.dta", clear

keep in 1/383

```
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Importanceagree.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Importance.dta", clear

keep in 384/766

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Importancedisagree.dta"

* Merge datasets - only need agree for analysis

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/Importanceagree.dta", clear

rename Row irow

rename RowSE irowse

rename B NSDUHssr

merge 1:1 ssr_n using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.01.24/Decisionagree.dta"

drop _merge

rename Row drow

rename RowSE drowse

rename B dssr

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/DCSNIMPTagree.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/DCSNIMPTagree.dta", clear
```

```

merge 1:1 ssr_n using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.01.24/Friendshipagree.dta"

drop _merge

rename Row frow

rename RowSE frowse

rename B fssr

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/DCSNIMPTFRNDagree.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/DCSNIMPTFRNDagree.dta", clear

merge 1:1 ssr_n using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.01.24/Attendanceagree.dta"

drop _merge

rename Row arow

rename RowSE arowse

rename B assr

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/allagree.dta"

drop RowCIlower RowCIupper WeightedCount CountSE dssr fssr assr

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.01.24/allagree.dta", replace

/* Edits made to STREG excel file to match CDC

Kusilvak, AK added 2:AK, 2:Northern

```

Prince of Wales Outer Ketchikan, AK renamed to Prince of Wales Hyder, AK (Outer Ketchikan  
-> Hyder change made with 2010 Census)

Broomfield, CO added 8:CO, 4:Regions 2 and 7

Kalawao added 15:HI, 4:Maui

Shannon, SD renamed to Oglala Lakota, SD (County renamed in 2015, CDC uses new name in  
documentation)

Bedford City, VA and Clifton Forge, VA removed

Data saved as STREGclean \*/

/\* Merge with NSDUH religion variable(s) \*/

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.01.24/allagree.dta"

drop if STATENUM == 6 & SSRNUM == 1 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 2 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 3 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 4 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 5 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 6 /\* LA \*/

drop if STATENUM == 6 & SSRNUM == 7 /\* LA \*/

drop if STATENUM == 9 /\* CT \*/

drop if STATENUM == 10 & SSRNUM == 2 /\* DE \*/

drop if STATENUM == 10 & SSRNUM == 4 /\* DE \*/

drop if STATENUM == 11 /\* DC \*/

drop if STATENUM == 25 /\* MA \*/

```

drop if STATENUM == 26 & SSRNUM == 1 /* MI */

* 354 obs. - matches above dataset

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.07.24/Re-do/agree_v1.dta"

rename NSDUHssr nsduhsubstateregion

drop K RELIGIOUSBELIEFSINFLUENCELIFE ITISIMPORTANTTHATMYFRIENDS
PAST12MOSHOWMANYRELIGSER MYRELIGIOUSBELIEFSAREVERYIM

order nsduhsubstateregion drow drowse irow irowse frow-rowse

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.07.24/Re-do/agree_merge.dta"

/* Aim 2 */

* Generate 2012-2019 file for analysis */

* Began using only 2015-2017 so was created first

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2015 CDC.xlsx", sheet("2015 CDC") firstrow

drop in 1

* Generate full state names *

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

```



replace state = "Connecticut" if C==" CT"  
replace state = "Delaware" if C==" DE"  
replace state = "Florida" if C==" FL"  
replace state = "Georgia" if C==" GA"  
replace state = "Hawaii" if C==" HI"  
replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA "  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA"  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"

replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"  
replace state = "Texas" if C==" TX"  
replace state = "Utah" if C==" UT"  
replace state = "Vermont" if C==" VT"  
replace state = "Virginia" if C==" VA"  
replace state = "Washington" if C==" WA"  
replace state = "West Virginia" if C==" WV"  
replace state = "Wisconsin" if C==" WI"  
replace state = "Wyoming" if C==" WY"

\* missing DC - need to add wards to doc \*

replace Geography = substr(Geography, "County", "", .)

```

replace Geography = substr(Geography, "Parish", "", .)
replace Geography = substr(Geography, "Municipality", "", .)
replace Geography = substr(Geography, "Census Area", "", .)
replace Geography = substr(Geography, "-", "", .)
replace Geography = substr(Geography, "Borough", "", .)
replace Geography = substr(Geography, " ", "", .)
replace Geography = substr(Geography, ".", "", .)
replace Geography = substr(Geography, "Cityand", "", .)
replace Geography = substr(Geography, "", "", .)
rename Geography County
rename state State

/* Changes in 2012-2019 year files for merge:

Baltimore City to BaltimoreCity

DeKalb to DeKalb

Desoto to DeSoto

Dupage to DuPage

Lagrange to LaGrange

Laporte to LaPorte

Lasalle to LaSalle

Changed capitalization for counties with "Mc..." ex. McClain ... McPherson

O'brien to OBrien

Ste. Genevieve to StGenevieve

*/

```

```

* Drop large areas with SSR overlap or un-separated data *
drop if County == "DistrictofColumbia"
drop if County == "LosAngeles"
* Drop counties with no data available - all appear to be in Alaska *
drop if County == "Chugach"
drop if County == "CopperRiver"
drop if County == "HoonahAngoon"
drop if County == "Petersburg"
drop if County == "PrinceofWalesOuterKetchikan"
drop if County == "Skagway"
drop if County == "Wrangell"
* Remove additional problem areas in anticipation of merge
order State County
drop if State == "Massachusetts"
drop if State == "Connecticut"
drop if State == "Delaware" & County == "NewCastle"
drop if State == "Michigan" & County == "Wayne"
* 3,114 rows
save "/Users/clairschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2015CDC.dta"
rename Chlamydia ChlamydiaCase15
rename E ChlamydiaRate15
rename EarlyNonPrimaryNonSecondary EarlySyphCase15

```

```

rename G EarlySyphRate15

rename Gonorrhea GonorrheaCase15

rename I GonorrheaRate15

rename PrimaryandSecondarySyphilis PrimSecSyphCase15

rename K PrimSecSyphRate15

rename TotalSyphilis totalsyph15

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/15CDC.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2016 CDC.xlsx", sheet("AtlasPlusTableData-15") firstrow clear

drop in 1

* Generate full state names *

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

replace state = "Connecticut" if C==" CT"

replace state = "Delaware" if C==" DE"

replace state = "Florida" if C==" FL"

replace state = "Georgia" if C==" GA"

replace state = "Hawaii" if C==" HI"

```

replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA "  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA "  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"

replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"  
replace state = "Texas" if C==" TX"  
replace state = "Utah" if C==" UT"  
replace state = "Vermont" if C==" VT"  
replace state = "Virginia" if C==" VA"  
replace state = "Washington" if C==" WA"  
replace state = "West Virginia" if C==" WV"  
replace state = "Wisconsin" if C==" WI"  
replace state = "Wyoming" if C==" WY"

\* missing DC - need to add wards to doc \*

replace Geography = substr(Geography, "County", "", .)  
replace Geography = substr(Geography, "Parish", "", .)  
replace Geography = substr(Geography, "Municipality", "", .)  
replace Geography = substr(Geography, "Census Area", "", .)  
replace Geography = substr(Geography, "-", "", .)  
replace Geography = substr(Geography, "Borough", "", .)

```

replace Geography = subinstr(Geography, " ", "", .)
replace Geography = subinstr(Geography, ".", "", .)
replace Geography = subinstr(Geography, "Cityand", "", .)
replace Geography = subinstr(Geography, "", "", .)
rename Geography County
rename state State

/* Changes in 2012-2019 year files for merge:

Baltimore City to BaltimoreCity

DeKalb to DeKalb

Desoto to DeSoto

Dupage to DuPage

Lagrange to LaGrange

Laporte to LaPorte

Lasalle to LaSalle

Changed capitalization for counties with "Mc..." ex. McClain ... McPherson

O'brien to OBrien

Ste. Genevieve to StGenevieve

*/

* Drop large areas with SSR overlap or un-separated data *

drop if County == "DistrictofColumbia"

drop if County == "LosAngeles"

* Drop counties with no data available - all appear to be in Alaska *

drop if County == "Chugach"

```



```
drop if County == "CopperRiver"
drop if County == "HoonahAngoon"
drop if County == "Petersburg"
drop if County == "PrinceofWalesOuterKetchikan"
drop if County == "Skagway"
drop if County == "Wrangell"
* Remove additional problem areas in anticipation of merge
order State County
drop if State == "Massachusetts"
drop if State == "Connecticut"
drop if State == "Delaware" & County == "NewCastle"
drop if State == "Michigan" & County == "Wayne"
* 3,114 rows
save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2016CDC.dta"
clear
use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/15CDC.dta"
merge 1:1 State County using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.19.24/2016CDC.dta"
drop L _merge
rename Chlamydia ChlamydiaCase16
rename E ChlamydiaRate16
```

```

rename EarlyNonPrimaryNonSecondary EarlySyphCase16
rename G EarlySyphRate16
rename Gonorrhea GonorrheaCase16
rename I GonorrheaRate16
rename PrimaryandSecondarySyphilis PrimSecSyphCase16
rename K PrimSecSyphRate16
rename TotalSyphilis totalsyph16
save "/Users/claeschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/1516CDC.dta"
import excel "/Users/claeschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2017 CDC.xlsx", sheet("AtlasPlusTableData-16") firstrow clear
drop in 1
* Generate full state names *
gen state = "Alaska" if C==" AK"
replace state = "Alabama" if C==" AL"
replace state = "Arizona" if C==" AZ"
replace state = "Arkansas" if C==" AR"
replace state = "California" if C==" CA"
replace state = "Colorado" if C==" CO"
replace state = "Connecticut" if C==" CT"
replace state = "Delaware" if C==" DE"
replace state = "Florida" if C==" FL"
replace state = "Georgia" if C==" GA"

```

replace state = "Hawaii" if C==" HI"  
replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA"  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA "  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"

replace state = "North Dakota" if C==" ND"

replace state = "Ohio" if C==" OH"

replace state = "Oklahoma" if C==" OK"

replace state = "Oregon" if C==" OR"

replace state = "Pennsylvania" if C==" PA"

replace state = "Rhode Island" if C==" RI"

replace state = "South Carolina" if C==" SC"

replace state = "South Dakota" if C==" SD"

replace state = "Tennessee" if C==" TN"

replace state = "Texas" if C==" TX"

replace state = "Utah" if C==" UT"

replace state = "Vermont" if C==" VT"

replace state = "Virginia" if C==" VA"

replace state = "Washington" if C==" WA"

replace state = "West Virginia" if C==" WV"

replace state = "Wisconsin" if C==" WI"

replace state = "Wyoming" if C==" WY"

\* missing DC - need to add wards to doc \*

replace Geography = substr(Geography, "County", "", .)

replace Geography = substr(Geography, "Parish", "", .)

replace Geography = substr(Geography, "Municipality", "", .)

replace Geography = substr(Geography, "Census Area", "", .)

replace Geography = substr(Geography, "-", "", .)

```

replace Geography = substr(Geography, "Borough", "", .)
replace Geography = substr(Geography, " ", "", .)
replace Geography = substr(Geography, ".", "", .)
replace Geography = substr(Geography, "Cityand", "", .)
replace Geography = substr(Geography, "", "", .)

rename Geography County

rename state State

/* Changes in 2012-2019 year files for merge:

Baltimore City to BaltimoreCity

DeKalb to DeKalb

Desoto to DeSoto

Dupage to DuPage

Lagrange to LaGrange

Laporte to LaPorte

Lasalle to LaSalle

Changed capitalization for counties with "Mc..." ex. McClain ... McPherson

O'brien to OBrien

Ste. Genevieve to StGenevieve

*/

* Drop large areas with SSR overlap or un-separated data *

drop if County == "DistrictofColumbia"

drop if County == "LosAngeles"

* Drop counties with no data available - all appear to be in Alaska *

```

```
drop if County == "Chugach"
drop if County == "CopperRiver"
drop if County == "HoonahAngoon"
drop if County == "Petersburg"
drop if County == "PrinceofWalesOuterKetchikan"
drop if County == "Skagway"
drop if County == "Wrangell"
* Remove additional problem areas in anticipation of merge
order State County
drop if State == "Massachusetts"
drop if State == "Connecticut"
drop if State == "Delaware" & County == "NewCastle"
drop if State == "Michigan" & County == "Wayne"
* 3,114 rows
save "/Users/clareshertzinger/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/2017CDC.dta"
clear
use "/Users/clareshertzinger/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/1516CDC_merge.dta"
merge 1:1 State County using "/Users/clareshertzinger/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.19.24/2017CDC.dta"
drop L _merge
rename Chlamydia ChlamydiaCase17
```

```
rename E ChlamydiaRate17
rename EarlyNonPrimaryNonSecondary EarlySyphCase17
rename G EarlySyphRate17
rename Gonorrhea GonorrheaCase17
rename I GonorrheaRate17
rename PrimaryandSecondarySyphilis PrimSecSyphCase17
rename K PrimSecSyphRate17
rename TotalSyphilis totalsyph17
save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/151617CDC.dta"
destring ChlamydiaCase15, replace
destring EarlySyphCase15, replace
destring GonorrheaCase15, replace
destring PrimSecSyphCase15, replace
destring ChlamydiaCase16, replace
destring EarlySyphCase16, replace
destring GonorrheaCase16, replace
destring PrimSecSyphCase16, replace
destring ChlamydiaCase17, replace
destring EarlySyphCase17, replace
destring GonorrheaCase17, replace
destring PrimSecSyphCase17, replace
```

```

drop Year ChlamydiaRate15 EarlySyphRate15 GonorrheaRate15 PrimSecSyphRate15
ChlamydiaRate16 EarlySyphRate16 GonorrheaRate16 PrimSecSyphRate16 ChlamydiaRate17
EarlySyphRate17 GonorrheaRate17 PrimSecSyphRate17
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/151617CDC_casesonly.dta"
* re-do STREG data to allow for merging
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.02.24/STREG Definitions 2000 clean.xlsx", sheet("ustracts2k") firstrow clear
keep State County NSDUHSubstateRegion
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/0207STREG.dta"
gen first =_n
by County NSDUHSubstateRegion (first), sort: generate order = _n == 1
drop first
keep if order == 1
sort State County
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/0219STREG.dta"
drop if County == "Los Angeles"
drop if County == "LosAngeles"
drop if State == "Connecticut"
drop if State == "Delaware" & County == "NewCastle"
drop if State == "District of Columbia"

```



```

drop if State == "Massachusetts"

drop if state == "Michigan" & county == "Wayne"

drop if state == "Alaska" & county == "WadeHampton"

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.19.24/0219STREG.dta", replace

merge 1:1 state county using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.20.24/151617CDC_casesonly.dta"

* successful merge, 3,114 rows

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.20.24/151617STDbycounty.dta"

/* import additional years to merge with existing 15-17 data set */

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/2012 STD.xlsx", sheet("AtlasPlusTableData-20") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2012CDC.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2013 STD.xlsx", sheet("AtlasPlusTableData-16") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2013CDC.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2014 STD.xlsx", sheet("AtlasPlusTableData-17") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2014CDC.dta"

```

```

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2018 STD.xlsx", sheet("AtlasPlusTableData-18") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2018CDC.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2019 STD.xlsx", sheet("AtlasPlusTableData-19") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2019CDC.dta"

/* 2012 */

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2012CDC.dta", clear

drop in 1

* Generate full state names *

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

replace state = "Connecticut" if C==" CT"

replace state = "Delaware" if C==" DE"

replace state = "Florida" if C==" FL"

replace state = "Georgia" if C==" GA"

```

replace state = "Hawaii" if C==" HI"  
replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA"  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA "  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"

replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"  
replace state = "Texas" if C==" TX"  
replace state = "Utah" if C==" UT"  
replace state = "Vermont" if C==" VT"  
replace state = "Virginia" if C==" VA"  
replace state = "Washington" if C==" WA"  
replace state = "West Virginia" if C==" WV"  
replace state = "Wisconsin" if C==" WI"  
replace state = "Wyoming" if C==" WY"  
replace Geography = substr(Geography, "County", "", .)  
replace Geography = substr(Geography, "Parish", "", .)  
replace Geography = substr(Geography, "Municipality", "", .)  
replace Geography = substr(Geography, "Census Area", "", .)  
replace Geography = substr(Geography, "-", "", .)  
replace Geography = substr(Geography, "Borough", "", .)

replace Geography = substr(Geography, " ", "", .)

replace Geography = substr(Geography, ".", "", .)

replace Geography = substr(Geography, "Cityand", "", .)

replace Geography = substr(Geography, "", "", .)

rename Geography County

rename state State

\* Drop large areas with SSR overlap or un-separated data \*

drop if County == "DistrictofColumbia"

drop if County == "LosAngeles"

\* Drop counties with no data available - all appear to be in Alaska \*

drop if County == "Chugach"

drop if County == "CopperRiver"

drop if County == "HoonahAngoon"

drop if County == "Petersburg"

drop if County == "PrinceofWalesOuterKetchikan"

drop if County == "Skagway"

drop if County == "Wrangell"

\* Remove additional problem areas in anticipation of merge

order State County

drop if State == "Massachusetts"

drop if State == "Connecticut"

drop if State == "Delaware" & County == "NewCastle"

drop if State == "Michigan" & County == "Wayne"

```
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2012CDC.dta", replace
rename Chlamydia ChlamydiaCase12
rename E ChlamydiaRate12
rename EarlyNonPrimaryNonSecondary EarlySyphCase12
rename G EarlySyphRate12
rename Gonorrhea GonorrheaCase12
rename I GonorrheaRate12
rename PrimaryandSecondarySyphilis PrimSecSyphCase12
rename K PrimSecSyphRate12
rename TotalSyphilis totalsyph12
drop L
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2012CDC.dta", replace
use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.07.24/0219STREG.dta", clear
merge 1:1 State County using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/2012CDC.dta"
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/12CDC.dta"
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2013 STD.xlsx", sheet("AtlasPlusTableData-16") firstrow clear
```

```
save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.28.24/2013CDC.dta"
```

```
drop in 1
```

```
* Generate full state names *
```

```
gen state = "Alaska" if C==" AK"
```

```
replace state = "Alabama" if C==" AL"
```

```
replace state = "Arizona" if C==" AZ"
```

```
replace state = "Arkansas" if C==" AR"
```

```
replace state = "California" if C==" CA"
```

```
replace state = "Colorado" if C==" CO"
```

```
replace state = "Connecticut" if C==" CT"
```

```
replace state = "Delaware" if C==" DE"
```

```
replace state = "Florida" if C==" FL"
```

```
replace state = "Georgia" if C==" GA"
```

```
replace state = "Hawaii" if C==" HI"
```

```
replace state = "Idaho" if C==" ID"
```

```
replace state = "Illinois" if C==" IL"
```

```
replace state = "Indiana" if C==" IN"
```

```
replace state = "Iowa" if C==" IA"
```

```
replace state = "Kansas" if C==" KS"
```

```
replace state = "Kentucky" if C==" KY"
```

```
replace state = "Louisiana" if C==" LA"
```

```
replace state = "Maine" if C==" ME"
```

replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA"  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"



```

replace state = "Texas" if C==" TX"

replace state = "Utah" if C==" UT"

replace state = "Vermont" if C==" VT"

replace state = "Virginia" if C==" VA"

replace state = "Washington" if C==" WA"

replace state = "West Virginia" if C==" WV"

replace state = "Wisconsin" if C==" WI"

replace state = "Wyoming" if C==" WY"

replace Geography = substr(Geography, "County", "", .)

replace Geography = substr(Geography, "Parish", "", .)

replace Geography = substr(Geography, "Municipality", "", .)

replace Geography = substr(Geography, "Census Area", "", .)

replace Geography = substr(Geography, "-", "", .)

replace Geography = substr(Geography, "Borough", "", .)

replace Geography = substr(Geography, " ", "", .)

replace Geography = substr(Geography, ".", "", .)

replace Geography = substr(Geography, "Cityand", "", .)

replace Geography = substr(Geography, "", "", .)

rename Geography County

rename state State

* Drop large areas with SSR overlap or un-separated data *

drop if County == "DistrictofColumbia"

drop if County == "LosAngeles"

```

\* Drop counties with no data available - all appear to be in Alaska \*

drop if County == "Chugach"

drop if County == "CopperRiver"

drop if County == "HoonahAngoon"

drop if County == "Petersburg"

drop if County == "PrinceofWalesOuterKetchikan"

drop if County == "Skagway"

drop if County == "Wrangell"

\* Remove additional problem areas in anticipation of merge

order State County

drop if State == "Massachusetts"

drop if State == "Connecticut"

drop if State == "Delaware" & County == "NewCastle"

drop if State == "Michigan" & County == "Wayne"

rename Chlamydia ChlamydiaCase13

rename E ChlamydiaRate13

rename EarlyNonPrimaryNonSecondary EarlySyphCase13

rename G EarlySyphRate13

rename Gonorrhea GonorrheaCase13

rename I GonorrheaRate13

rename PrimaryandSecondarySyphilis PrimSecSyphCase13

rename K PrimSecSyphRate13

rename TotalSyphilis totalsyph13

```

drop L

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2013CDC.dta", replace

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/12CDC.dta", clear

drop _merge

merge 1:1 State County using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/2013CDC.dta"

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1213CDC.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2014 STD.xlsx", sheet("AtlasPlusTableData-17") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2014CDC.dta"

drop in 1

* Generate full state names *

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

replace state = "Connecticut" if C==" CT"

```

replace state = "Delaware" if C==" DE"  
replace state = "Florida" if C==" FL"  
replace state = "Georgia" if C==" GA"  
replace state = "Hawaii" if C==" HI"  
replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA "  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA "  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"

replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"  
replace state = "Texas" if C==" TX"  
replace state = "Utah" if C==" UT"  
replace state = "Vermont" if C==" VT"  
replace state = "Virginia" if C==" VA"  
replace state = "Washington" if C==" WA"  
replace state = "West Virginia" if C==" WV"  
replace state = "Wisconsin" if C==" WI"  
replace state = "Wyoming" if C==" WY"  
replace Geography = substr(Geography, "County", "", .)  
replace Geography = substr(Geography, "Parish", "", .)  
replace Geography = substr(Geography, "Municipality", "", .)

replace Geography = substr(Geography, "Census Area", "", .)

replace Geography = substr(Geography, "-", "", .)

replace Geography = substr(Geography, "Borough", "", .)

replace Geography = substr(Geography, " ", "", .)

replace Geography = substr(Geography, ".", "", .)

replace Geography = substr(Geography, "Cityand", "", .)

replace Geography = substr(Geography, "", "", .)

rename Geography County

rename state State

\* Drop large areas with SSR overlap or un-separated data \*

drop if County == "DistrictofColumbia"

drop if County == "LosAngeles"

\* Drop counties with no data available - all appear to be in Alaska \*

drop if County == "Chugach"

drop if County == "CopperRiver"

drop if County == "HoonahAngoon"

drop if County == "Petersburg"

drop if County == "PrinceofWalesOuterKetchikan"

drop if County == "Skagway"

drop if County == "Wrangell"

\* Remove additional problem areas in anticipation of merge

order State County

drop if State == "Massachusetts"

```
drop if State == "Connecticut"

drop if State == "Delaware" & County == "NewCastle"

drop if State == "Michigan" & County == "Wayne"

rename Chlamydia ChlamydiaCase14

rename E ChlamydiaRate14

rename EarlyNonPrimaryNonSecondary EarlySyphCase14

rename G EarlySyphRate14

rename Gonorrhea GonorrheaCase14

rename I GonorrheaRate14

rename PrimaryandSecondarySyphilis PrimSecSyphCase14

rename K PrimSecSyphRate14

rename TotalSyphilis totalsyph14

drop L

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2014CDC.dta", replace

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1213CDC.dta", clear

drop _merge

merge 1:1 State County using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/2014CDC.dta"

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/121314CDC.dta"
```

```

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/121314CDC.dta", replace

merge 1:1 State County using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.20.24/151617_STDbySSR.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/121314151617CDC.dta"

/* 2018 */

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2018 STD.xlsx", sheet("AtlasPlusTableData-18") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2018CDC.dta", replace

drop in 1

* Generate full state names *

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

replace state = "Connecticut" if C==" CT"

replace state = "Delaware" if C==" DE"

```



replace state = "Florida" if C==" FL"  
replace state = "Georgia" if C==" GA"  
replace state = "Hawaii" if C==" HI"  
replace state = "Idaho" if C==" ID"  
replace state = "Illinois" if C==" IL"  
replace state = "Indiana" if C==" IN"  
replace state = "Iowa" if C==" IA"  
replace state = "Kansas" if C==" KS"  
replace state = "Kentucky" if C==" KY"  
replace state = "Louisiana" if C==" LA "  
replace state = "Maine" if C==" ME"  
replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA "  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"

replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"  
replace state = "Texas" if C==" TX"  
replace state = "Utah" if C==" UT"  
replace state = "Vermont" if C==" VT"  
replace state = "Virginia" if C==" VA"  
replace state = "Washington" if C==" WA"  
replace state = "West Virginia" if C==" WV"  
replace state = "Wisconsin" if C==" WI"  
replace state = "Wyoming" if C==" WY"  
replace Geography = substr(Geography, "County", "", .)  
replace Geography = substr(Geography, "Parish", "", .)  
replace Geography = substr(Geography, "Municipality", "", .)  
replace Geography = substr(Geography, "Census Area", "", .)

```

replace Geography = substr(Geography, "-", "", .)
replace Geography = substr(Geography, "Borough", "", .)
replace Geography = substr(Geography, " ", "", .)
replace Geography = substr(Geography, ".", "", .)
replace Geography = substr(Geography, "Cityand", "", .)
replace Geography = substr(Geography, "", "", .)
rename Geography County
rename state State

* Drop large areas with SSR overlap or un-separated data *
drop if County == "DistrictofColumbia"
drop if County == "LosAngeles"

* Drop counties with no data available - all appear to be in Alaska *
drop if County == "Chugach"
drop if County == "CopperRiver"
drop if County == "HoonahAngoon"
drop if County == "Petersburg"
drop if County == "PrinceofWalesOuterKetchikan"
drop if County == "Skagway"
drop if County == "Wrangell"

* Remove additional problem areas in anticipation of merge
order State County
drop if State == "Massachusetts"
drop if State == "Connecticut"

```

```

drop if State == "Delaware" & County == "NewCastle"

drop if State == "Michigan" & County == "Wayne"

rename Chlamydia ChlamydiaCase18

rename E ChlamydiaRate18

rename EarlyNonPrimaryNonSecondary EarlySyphCase18

rename G EarlySyphRate18

rename Gonorrhea GonorrheaCase18

rename I GonorrheaRate18

rename PrimaryandSecondarySyphilis PrimSecSyphCase18

rename K PrimSecSyphRate18

rename TotalSyphilis totalsyph18

drop L

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2018CDC.dta", replace

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/121314151617CDC.dta", clear

merge 1:1 State County using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/2018CDC.dta"

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/12131415161718CDC.dta"

import excel "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2019 STD.xlsx", sheet("AtlasPlusTableData-19") firstrow clear

```

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/02.28.24/2019CDC.dta", replace

drop in 1

\* Generate full state names \*

gen state = "Alaska" if C==" AK"

replace state = "Alabama" if C==" AL"

replace state = "Arizona" if C==" AZ"

replace state = "Arkansas" if C==" AR"

replace state = "California" if C==" CA"

replace state = "Colorado" if C==" CO"

replace state = "Connecticut" if C==" CT"

replace state = "Delaware" if C==" DE"

replace state = "Florida" if C==" FL"

replace state = "Georgia" if C==" GA"

replace state = "Hawaii" if C==" HI"

replace state = "Idaho" if C==" ID"

replace state = "Illinois" if C==" IL"

replace state = "Indiana" if C==" IN"

replace state = "Iowa" if C==" IA"

replace state = "Kansas" if C==" KS"

replace state = "Kentucky" if C==" KY"

replace state = "Louisiana" if C==" LA"

replace state = "Maine" if C==" ME"

replace state = "Maryland" if C==" MD"  
replace state = "Massachusetts" if C==" MA"  
replace state = "Michigan" if C==" MI"  
replace state = "Minnesota" if C==" MN"  
replace state = "Mississippi" if C==" MS"  
replace state = "Missouri" if C==" MO"  
replace state = "Montana" if C==" MT"  
replace state = "Nebraska" if C==" NE"  
replace state = "Nevada" if C==" NV"  
replace state = "New Hampshire" if C==" NH"  
replace state = "New Jersey" if C==" NJ"  
replace state = "New Mexico" if C==" NM"  
replace state = "New York" if C==" NY"  
replace state = "North Carolina" if C==" NC"  
replace state = "North Dakota" if C==" ND"  
replace state = "Ohio" if C==" OH"  
replace state = "Oklahoma" if C==" OK"  
replace state = "Oregon" if C==" OR"  
replace state = "Pennsylvania" if C==" PA"  
replace state = "Rhode Island" if C==" RI"  
replace state = "South Carolina" if C==" SC"  
replace state = "South Dakota" if C==" SD"  
replace state = "Tennessee" if C==" TN"

```

replace state = "Texas" if C==" TX"
replace state = "Utah" if C==" UT"
replace state = "Vermont" if C==" VT"
replace state = "Virginia" if C==" VA"
replace state = "Washington" if C==" WA"
replace state = "West Virginia" if C==" WV"
replace state = "Wisconsin" if C==" WI"
replace state = "Wyoming" if C==" WY"
replace Geography = substr(Geography, "County", "", .)
replace Geography = substr(Geography, "Parish", "", .)
replace Geography = substr(Geography, "Municipality", "", .)
replace Geography = substr(Geography, "Census Area", "", .)
replace Geography = substr(Geography, "-", "", .)
replace Geography = substr(Geography, "Borough", "", .)
replace Geography = substr(Geography, " ", "", .)
replace Geography = substr(Geography, ".", "", .)
replace Geography = substr(Geography, "Cityand", "", .)
replace Geography = substr(Geography, "", "", .)
rename Geography County
rename state State

```

\* Drop large areas with SSR overlap or un-separated data \*

```
drop if County == "DistrictofColumbia"
```

drop if County == "LosAngeles"

\* Drop counties with no data available - all appear to be in Alaska \*

drop if County == "Chugach"

drop if County == "CopperRiver"

drop if County == "HoonahAngoon"

drop if County == "Petersburg"

drop if County == "PrinceofWalesOuterKetchikan"

drop if County == "Skagway"

drop if County == "Wrangell"

\* Remove additional problem areas in anticipation of merge

order State County

drop if State == "Massachusetts"

drop if State == "Connecticut"

drop if State == "Delaware" & County == "NewCastle"

drop if State == "Michigan" & County == "Wayne"

rename Chlamydia ChlamydiaCase19

rename E ChlamydiaRate19

rename EarlyNonPrimaryNonSecondary EarlySyphCase19

rename G EarlySyphRate19

rename Gonorrhea GonorrheaCase19

rename I GonorrheaRate19

rename PrimaryandSecondarySyphilis PrimSecSyphCase19

rename K PrimSecSyphRate19



```
rename TotalSyphilis totalsyph19

drop L

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/2019CDC.dta", replace

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/12131415161718CDC.dta", clear

drop _merge

merge 1:1 State County using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/2019CDC.dta"

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1213141516171819CDC.dta"

drop FIPS Year C _merge

order NSDUHSubstateRegion State County

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1213141516171819CDC.dta", replace

/* Destring to allow for re-formatting */

destring ChlamydiaCase12, replace

destring EarlySyphCase12, replace

destring GonorrheaCase12, replace

destring PrimSecSyphCase12, replace

destring ChlamydiaCase13, replace

destring EarlySyphCase13, replace
```

destring GonorrheaCase13, replace  
destring PrimSecSyphCase13, replace  
destring ChlamydiaCase14, replace  
destring EarlySyphCase14, replace  
destring GonorrheaCase14, replace  
destring PrimSecSyphCase14, replace  
destring ChlamydiaCase18, replace  
destring EarlySyphCase18, replace  
destring GonorrheaCase18, replace  
destring PrimSecSyphCase18, replace  
destring ChlamydiaCase19, replace  
destring EarlySyphCase19, replace  
destring GonorrheaCase19, replace  
destring PrimSecSyphCase19, replace

\* for sake of ease/number of values drop rates

drop ChlamydiaRate12 EarlySyphRate12 GonorrheaRate12 PrimSecSyphRate12  
ChlamydiaRate13 EarlySyphRate13 GonorrheaRate13 PrimSecSyphRate13 ChlamydiaRate14  
EarlySyphRate14 GonorrheaRate14 PrimSecSyphRate14 ChlamydiaRate15 EarlySyphRate15  
GonorrheaRate15 PrimSecSyphRate15 ChlamydiaRate16 EarlySyphRate16 GonorrheaRate16  
PrimSecSyphRate16 ChlamydiaRate17 EarlySyphRate17 GonorrheaRate17 PrimSecSyphRate17  
ChlamydiaRate18 EarlySyphRate18 GonorrheaRate18 PrimSecSyphRate18 ChlamydiaRate19  
EarlySyphRate19 GonorrheaRate19 PrimSecSyphRate19  
rename \*, lower

```

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1213141516171819CDC.dta", replace

merge 1:1 state county using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.19.24/0219STREG.dta"

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1219CDCbyCounty.dta"

```

\*I want to create a numeric county id to represent as my timepoints for when I have to reshape the panel data format to wide format

```

sort nsduhsubstateregion county

by nsduhsubstateregion: gen countyid = _n

order countyid

```

\*Create globals for the macros

```

global disease "chlamydiacase gonorrheacase totalsyph"

global years "12 13 14 15 16 17 18 19"

```

\*\*\*I need to rename the variable names so when I convert to wide format, the variable names will be 02\_1, 02\_2, etc instead of 021, 022, etc

```

    foreach i of global disease{
        foreach n of global years{
            rename `i`n' `i`n'_
        }
    }

```

```

keep nsduhsubstateregion countyid chlamydiacase* gonorrheacase* totalsyph*

```

\*Reshape the long format to wide format

```
reshape wide chlamydiacase* gonorrheacase* totalsyph*, i(nsduhsubstateregion) j(countyid)
```

```
* yes this worked, ok proceed
```

```
*Sum all the cases from different counties for each year into a single variable
```

```
foreach d of global disease{
```

```
    foreach n of global years{
```

```
        egen `d`n' = rowtotal(`d`n'*)
```

```
    }
```

```
}
```

```
*This is optional. I want to get rid of the unnecessary variables as I already used them to sum all cases into a single variable.
```

```
foreach i of global disease{
```

```
    foreach n of global years{
```

```
        drop `i`n'*
```

```
    }
```

```
}
```

```
order nsduhsubstateregion chlamydiacase_* gonorrheacase_* totalsyph*
```

```
*** gives error chlamydiacase_* not found but it works fine so ignore
```

```
save "/Users/claireschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.28.24/1219CDCbySSR.dta"
```

```
* this is cleaned for problematic SSR so pull from existing data set
```

```
use "/Users/claireschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/02.07.24/Re-do/agree_merge.dta"
```

```

keep nsduhsubstateregion STATENUM SSRNUM ssr_n irow irowse drow drowse frow frowse
arow arowse

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/NSDUHonly.dta"

merge 1:1 nsduhsubstateregion using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/02.28.24/1219CDCbySSR.dta"

* perfect merge with 354 SSR

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/1219CDCNSDUH.dta"

egen chlamydiatotal=rowtotal(chlamydiacase12-chlamydiacase19)

egen gonorrheatotal=rowtotal(gonorrheacase12-gonorrheacase19)

egen syphilistotal=rowtotal(totalsyph12-totalsyph19)

egen STItotal=rowtotal(chlamydiacase12-totalsyph19)

***STItotal

drop _merge

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/totalCDCNSDUH.dta"

/* Aim 1 */

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/totalCDCNSDUH.dta", clear

* fairly certain I need to add in youth or recode this somehow but unclear how

summarize irow drow frow arow

factor drow irow frow arow

```

```

sem (drow irow frow arow <- Religiosity), cov(e.irow*e.arow)

predict Relighat, latent

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/totalCDCNSDUH.dta", replace

/* Aim 2 */

* add age distribution to the data set

import excel "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/CATAG3 x STREG10.xlsx", sheet("CATAG3 x STREG10") firstrow clear

rename AGECATEGORYRECODE5LEVELS agecat

rename B nsduhsubstateregion

rename Row AgeRow

rename RowSE AgeRowSE

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Age.dta"

drop if nsduhsubstateregion == "Overall"

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/02.28.24/totalCDCNSDUH.dta", clear

drop _merge

merge 1:m nsduhsubstateregion using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Age.dta"

drop if chlamydiacase12 == .

* removes problematic areas

```

```

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/03.20.24/1219CDCNSDUH_agedist.dta"

encode agecat, gen(agecatn)

order nsduhsubstateregion stssrn agecatn AgeRow

sort stssrn agecatn

keep nsduhsubstateregion stssrn agecatn AgeRow

reshape wide AgeRow, i(stssrn) j(agecatn)

rename AgeRow1 pct1217

rename AgeRow2 pct1825

rename AgeRow3 pct2635

rename AgeRow4 pct3649

rename AgeRow5 pct50up

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/03.22.24/agecat.dta"

gsem (drow irow frow arow <- Religiosity), cov(e.irow*e.arow)

predict rhat, latent

*Bayesian estimate

*re-run from above

sem (drow irow frow arow <- Religiosity), cov(e.irow*e.arow)

predict Relighat, latent

*Factor score

corr Relighat rhat

```

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/1219Age.dta"

\*test poisson vs negative binomial model

poisson STItotal Relighat

estat gof

nbreg STItotal Relighat

sum STItotal, detail

nbreg chlamydiatotal Relighat

nbreg gonorrhheatotal Relighat

nbreg syphilistotal Relighat

\*test truncated nbreg for data with no zeros

tnbreg chlamydiatotal Relighat

nbreg chlamydiatotal Relighat

tnbreg gonorrhheatotal Relighat

nbreg gonorrhheatotal Relighat

tnbreg syphilistotal Relighat

nbreg syphilistotal Relighat

tnbreg STItotal Relighat

nbreg STItotal Relighat

\*all provide basically the same or exactly the same answers

\*get effect of population size on estimates

nbreg STItotal Relighat zpop

nbreg chlamydiatotal Relighat zpop



```

nbreg gonorrheatotal Relighat zpop
nbreg syphilistotal Relighat zpop
*get effect of age on estimates
nbreg STItotal Relighat zpop pct1217 pct1825 pct2635 pct3649
nbreg chlamydiatotal Relighat zpop pct1217 pct1825 pct2635 pct3649
nbreg gonorrheatotal Relighat zpop pct1217 pct1825 pct2635 pct3649
nbreg syphilistotal Relighat zpop pct1217 pct1825 pct2635 pct3649
/* Aim 3 */
*getting additional variables together
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct White.xlsx", sheet("Pct White") firstrow clear
rename B nsduhsubstateregion
rename Row WhiteRow
rename RowSE WhiteRowSE
drop RowCIlower RowCIupper
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhite.dta"
import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Hispanic.xlsx", sheet("Hispanic") firstrow clear
rename B nsduhsubstateregion
rename Row HispanicRow
rename RowSE HispanicRowSE
drop RowCIlower RowCIupper

```

```
save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pcthispanic.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhite.dta"

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisp.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Poverty.xlsx", sheet("Poverty") firstrow clear

rename B nsduhsubstateregion

rename Row PovertyRow

rename RowSE PovertyRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctpoverty.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisp.dta"

drop _merge

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctpoverty.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisppov.dta"
```

```

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct YrPsychotherapeuticUse.xlsx", sheet("YrPsychotherapeuticUse")

firstrow clear

rename B nsduhsubstateregion

rename Row PsychRow

rename RowSE PsychRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctpsych.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhitehisppov.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisppovpsych.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct YrDrugAbuse.xlsx", sheet("YrDrugAbuse") firstrow clear

rename B nsduhsubstateregion

rename Row DrugAbRow

rename RowSE DrugAbRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctdrugab.dta"

```

```

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhitehisppovpsych.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisppovpsychdrugab.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct YrAlcoholAbuse.xlsx", sheet("YrAlcoholAbuse") firstrow clear

rename B nsduhsubstateregion

rename Row AlcAbRow

rename RowSE AlcAbRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctalcab.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhitehisppovpsychdrugab.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisppovpsychdrugabalcab.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Uninsured.xlsx", sheet("IRINSUR4 x STREG10") firstrow clear

rename B nsduhsubstateregion

rename Row UninsuredRow

rename RowSE UninsuredRowSE

```

```

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctuninsured.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhitehisppovpsychdrugabalcab.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctwhitehisppovpsychdrugabalcabuninsured.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct EmploymentStatus.xlsx", sheet("EmploymentStatus") firstrow clear

rename B nsduhsubstateregion

rename Row EmployedRow

rename RowSE EmployedRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctemployed.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/pctwhitehisppovpsychdrugabalcabuninsured.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars.dta"

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct LTHighSchool.xlsx", sheet("Pct LTHighSchool") firstrow clear

```

```

rename B nsduhsubstateregion

rename Row EducRow

rename RowSE EducRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctlhighschool.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3vars.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars.dta", replace

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Male.xlsx", sheet("Sheet 1 - IRSEX x STREG10") firstrow clear

rename B nsduhsubstateregion

rename Row MaleRow

rename RowSE MaleRowSE

drop RowCIlower RowCIupper

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctmale.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3vars.dta"

drop _merge

```

```

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars.dta", replace

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Black.xlsx", sheet("Sheet1") firstrow clear

rename B nsduhsubstateregion

rename PropBlack BlackRow

keep nsduhsubstateregion BlackRow

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/pctBlack.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3data.dta"

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta", replace

* Added later

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Booked.xlsx", sheet("EverBooked") firstrow clear

rename B nsduhsubstateregion

rename Row BookedRow

rename RowSE BookedRowSE

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/Booked.dta"

```

```

merge 1:1 nsduhsubstateregion using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3vars.dta"

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars_v2.dta"

import excel "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct Tobacco 30 day.xlsx", sheet("Tobacco30day") firstrow clear

rename B nsduhsubstateregion

rename Row SmokeRow

rename RowSE SmokeRowSE

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/Smoker.dta"

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars_v2.dta"

drop _merge

merge 1:1 nsduhsubstateregion using "/Users/clareshertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/Smoker.dta"

drop _merge

keep nsduhsubstateregion BookedRow BookedRowSE SmokeRow SmokeRowSE

save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars_v2.dta", replace

*clean data to remove SSRs that cannot be matched (383 -> 354)

use "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/1219Age.dta", clear

```



```

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3vars.dta"

*because 1219Age is already cleaned merge returns 29 unmatched rows which can be dropped to
get final clean data set

keep in 1/354

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta"

*from later addition of booked and tobacco

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3vars_v2.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/aim3data.dta"

drop if irow == .

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta", replace

drop _merge

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta", replace

*add substate region population information into the data set

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data
files/Aim 3 Variables/Pct population.xlsx", sheet("STREG10") firstrow clear

rename A nsduhsubstateregion

```

```

sum WeightedCount

gen zpop = (WeightedCount-643148.8)/584445

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/Population.dta"

use "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta"

merge 1:1 nsduhsubstateregion using "/Users/claieschertzing/Documents/MSU Epidemiology
PhD/Dissertation/Data files/Aim 3 Variables/Population.dta"

keep in 1/354

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta", replace

drop _merge

order nsduhsubstateregion irow-totalsyph19 chlamydiatotal gonorrheatotal syphilistotal STItotal
Relighat pct1217-zpop

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3
Variables/aim3data.dta", replace

*Individual variable tests

nbreg STItotal Relighat

est store A0

lrtest A0 A

*holding constant pop. size

nbreg STItotal Relighat zpop

est store A

```

nbreg STItotal Relighat zpop pct1217

est store B

lrtest A B

nbreg STItotal Relighat zpop pct1217 pct1825

est store C

lrtest A C

nbreg STItotal Relighat zpop pct1217 pct1825 pct2635

est store D

lrtest A D

nbreg STItotal Relighat zpop pct1217 pct1825 pct2635 pct3649

est store E

lrtest A E

\*when 50+ is added model is misspecified (p-values are crazy) so leave out (high correlations between vars., especially for 50up group)

\*aim 3 subsec. show in age control section

\*show lrtests that contrast each group in a table

nbreg STItotal Relighat zpop WhiteRow HispanicRow

est store F

lrtest A F

\*another subsec. holding race constant (leave Black out to prevent misspecification)

nbreg STItotal Relighat zpop MaleRow

est store G

lrtest A G

\*holding gender dist. constant

nbreg STItotal Relighat zpop PovertyRow

est store H

lrtest A H

\*holding prop. of poverty constant

nbreg STItotal Relighat zpop PsychRow

est store I

lrtest A I

\*holding psychotherapeutic usage constant

nbreg STItotal Relighat zpop DrugAbRow

est store J

lrtest A J

\*holding drug use disorder constant

nbreg STItotal Relighat zpop AlcAbRow

est store K

lrtest A K

\*holding alcohol use disorder constant

nbreg STItotal Relighat zpop UninsuredRow

est store L

lrtest A L

\*holding prop. uninsured constant

nbreg STItotal Relighat zpop EmployedRow

est store M

lrtest A M

\*holding fully employed constant --> not significant

nbreg STItotal Relighat zpop EducRow

est store N

lrtest A N

\*holding constant education

nbreg STItotal Relighat zpop BookedRow

est store O

lrtest A O

\*keeping having been booked constant --> not significant

nbreg STItotal Relighat zpop SmokeRow

est store P

lrtest A P

\*keeping tobacco usage constant --> not significant

\* Do same for Chlamydia

nbreg chlamydiatotal Relighat

est store A0

nbreg chlamydiatotal Relighat zpop

est store A

lrtest A0 A

nbreg chlamydiatotal Relighat zpop pct1217 pct1825 pct2635 pct3649

est store B

lrtest A B

nbreg chlamydiatotal Relighat zpop WhiteRow HispanicRow

est store C

lrtest A C

nbreg chlamydiatotal Relighat zpop MaleRow

est store D

lrtest A D

nbreg chlamydiatotal Relighat zpop AlcAbRow

est store E

lrtest A E

nbreg chlamydiatotal Relighat zpop DrugAbRow

est store F

lrtest A F

nbreg chlamydiatotal Relighat zpop PovertyRow

est store G

lrtest A G

nbreg chlamydiatotal Relighat zpop EducRow

est store H

lrtest A H

nbreg chlamydiatotal Relighat zpop EmployedRow

est store I

lrtest A I

nbreg chlamydiatotal Relighat zpop UninsuredRow

est store J

lrtest A J

nbreg chlamydiatotal Relighat zpop BookedRow

est store K

lrtest A K

nbreg chlamydiatotal Relighat zpop SmokeRow

est store L

lrtest A L

nbreg chlamydiatotal Relighat zpop PsychRow

est store M

lrtest A M

\*Age-specific

nbreg chlamydiatotal Relighat zpop

est store A

nbreg chlamydiatotal Relighat zpop pct1217

est store B

lrtest A B

nbreg chlamydiatotal Relighat zpop pct1217 pct1825

est store C

lrtest A C

nbreg chlamydiatotal Relighat zpop pct1217 pct1825 pct2635

est store D

lrtest A D

\*again for gonorrhoea

nbreg gonorrheatotal Relighat

est store A0

nbreg gonorrheatotal Relighat zpop

est store A

lrtest A0 A

\*age-specific

nbreg gonorrheatotal Relighat zpop

est store A

nbreg gonorrheatotal Relighat zpop pct1217

est store B

lrtest A B

nbreg gonorrheatotal Relighat zpop pct1217 pct1825

est store C

lrtest A C

nbreg gonorrheatotal Relighat zpop pct1217 pct1825 pct2635

est store D

lrtest A D

\*now all together

nbreg gonorrheatotal Relighat zpop pct1217 pct1825 pct2635 pct3649

est store B

lrtest A B

nbreg gonorrheatotal Relighat zpop WhiteRow HispanicRow

est store C



lrtest A C

nbreg gonorrheatotal Relighat zpop MaleRow

est store D

lrtest A D

nbreg gonorrheatotal Relighat zpop AlcAbRow

est store E

lrtest A E

nbreg gonorrheatotal Relighat zpop DrugAbRow

est store F

lrtest A F

nbreg gonorrheatotal Relighat zpop PovertyRow

est store G

lrtest A G

nbreg gonorrheatotal Relighat zpop EducRow

est store H

lrtest A H

nbreg gonorrheatotal Relighat zpop EmployedRow

est store I

lrtest A I

nbreg gonorrheatotal Relighat zpop UninsuredRow

est store J

lrtest A J

nbreg gonorrheatotal Relighat zpop BookedRow

est store K

lrtest A K

nbreg gonorrhoeatotal Relighat zpop SmokeRow

est store L

lrtest A L

nbreg gonorrhoeatotal Relighat zpop PsychRow

est store M

lrtest A M

\*last for syphilis

nbreg syphilistotal Relighat

est store A0

nbreg syphilistotal Relighat zpop

est store A

lrtest A0 A

\*age-specific

nbreg syphilistotal Relighat zpop

est store A

nbreg syphilistotal Relighat zpop pct1217

est store B

lrtest A B

nbreg syphilistotal Relighat zpop pct1217 pct1825

est store C

lrtest A C

nbreg syphilistotal Relighat zpop pct1217 pct1825 pct2635

est store D

lrtest A D

\*now all together

nbreg syphilistotal Relighat zpop pct1217 pct1825 pct2635 pct3649

est store B

lrtest A B

nbreg syphilistotal Relighat zpop WhiteRow HispanicRow

est store C

lrtest A C

nbreg syphilistotal Relighat zpop MaleRow

est store D

lrtest A D

nbreg syphilistotal Relighat zpop AlcAbRow

est store E

lrtest A E

nbreg syphilistotal Relighat zpop DrugAbRow

est store F

lrtest A F

nbreg syphilistotal Relighat zpop PovertyRow

est store G

lrtest A G

nbreg syphilistotal Relighat zpop EducRow

est store H

lrtest A H

nbreg syphilistotal Relighat zpop EmployedRow

est store I

lrtest A I

nbreg syphilistotal Relighat zpop UninsuredRow

est store J

lrtest A J

nbreg syphilistotal Relighat zpop BookedRow

est store K

lrtest A K

nbreg syphilistotal Relighat zpop SmokeRow

est store L

lrtest A L

nbreg syphilistotal Relighat zpop PsychRow

est store M

lrtest A M

\*Aim 3 step 1 models with probable predictors

nbreg STItotal Relighat MaleRow EducRow AlcAbRow DrugAbRow HispanicRow WhiteRow

PovertyRow pct1217 pct1825 pct2635 pct3649 zpop

nbreg chlamydiatotal Relighat MaleRow EducRow AlcAbRow DrugAbRow HispanicRow

WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop

nbreg gonorrhoeatotal Relighat MaleRow EducRow AlcAbRow DrugAbRow HispanicRow  
WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop

nbreg syphilistotal Relighat MaleRow EducRow AlcAbRow DrugAbRow HispanicRow  
WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop

\*Aim 3 step 2 models with unlikely predictors

nbreg STItotal Relighat MaleRow DrugAbRow HispanicRow WhiteRow pct2635 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

nbreg chlamydiatotal Relighat MaleRow DrugAbRow HispanicRow WhiteRow pct2635 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

nbreg gonorrhoeatotal Relighat MaleRow DrugAbRow HispanicRow WhiteRow pct2635 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

nbreg syphilistotal DrugAbRow HispanicRow WhiteRow pct1217 pct2635 zpop BookedRow  
SmokeRow EmployedRow UninsuredRow PsychRow

\*Aim 3 step 3 best models

nbreg STItotal Relighat MaleRow HispanicRow WhiteRow pct2635 zpop EmployedRow  
UninsuredRow PsychRow

nbreg chlamydiatotal Relighat MaleRow HispanicRow WhiteRow pct2635 zpop EmployedRow  
UninsuredRow PsychRow

nbreg gonorrhoeatotal Relighat HispanicRow WhiteRow pct2635 zpop SmokeRow  
UninsuredRow PsychRow

nbreg syphilistotal HispanicRow WhiteRow pct1217 pct2635 zpop BookedRow PsychRow

\*Do I get the same significant covariates if I run a backwards stepwise regression?

stepwise, pe(0.05) pr(0.1): nbreg STItotal Relighat MaleRow EducRow AlcAbRow DrugAbRow  
HispanicRow WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop BookedRow  
SmokeRow EmployedRow UninsuredRow PsychRow

\*no difference, smoking included but  $p=0.052$  so insignificant

stepwise, pe(0.05) pr(0.1): nbreg chlamydiatotal Relighat MaleRow EducRow AlcAbRow  
DrugAbRow HispanicRow WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

\*no difference

stepwise, pe(0.05) pr(0.1): nbreg gonorrhoeatotal Relighat MaleRow EducRow AlcAbRow  
DrugAbRow HispanicRow WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

\*no difference, male and employed included but  $p=0.062$  and  $0.055$  respectively so insignificant

stepwise, pe(0.05) pr(0.1): nbreg syphilistotal Relighat MaleRow EducRow AlcAbRow  
DrugAbRow HispanicRow WhiteRow PovertyRow pct1217 pct1825 pct2635 pct3649 zpop  
BookedRow SmokeRow EmployedRow UninsuredRow PsychRow

\*post-estimation EDA

```
import excel "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data  
files/Aim 3 Variables/STD meta.xlsx", sheet("Sheet1") firstrow clear  
save "/Users/clareshertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3  
Variables/STDmeta.dta"  
  
drop CI  
  
meta set Slope SE  
  
meta sum
```

meta forestplot

graph save "Graph" "/Users/claieschertzing/Documents/MSU Epidemiology

PhD/Dissertation/Data files/Aim 3 Variables/STI forest.gph"

graph export "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3 Variables/STI forest.pdf", as(pdf) name("Graph")

import excel "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3 Variables/Chlamydia meta.xlsx", sheet("Sheet1") firstrow clear

save "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3 Variables/Chlamydiameta.dta"

meta set Slope SE

meta forestplot

graph save "Graph" "/Users/claieschertzing/Documents/MSU Epidemiology

PhD/Dissertation/Data files/Aim 3 Variables/Chlamydia forest.gph"

graph export "/Users/claieschertzing/Documents/MSU Epidemiology PhD/Dissertation/Data files/Aim 3 Variables/Chlamydia forest.pdf", as(pdf) name("Graph")