2017

The Molecular Epidemiology of Tuberculosis in South Carolina, 2005-2011: Estimates of Recent Transmission and Risk Factors for Genotype Clustering

Amy Kathleen Roach
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Amy Roach

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Ji Shen, Committee Chairperson, Public Health Faculty
Dr. Amy Wilson, Committee Member, Public Health Faculty
Dr. Stephen Nkansah-Amankra, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2017

Abstract

The Molecular Epidemiology of Tuberculosis in South Carolina, 2005-2011: Estimates of Recent Transmission and Risk Factors for Genotype Clustering

by

Amy K. Roach

MSPH, University of South Carolina, 2003

BS, University of South Carolina, 1998

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2017

Abstract

Because tuberculosis (TB) is a public health threat that continues to elude elimination in the United States, there is a need to identify contributing factors that may have implications for targeted control measures. Molecular studies of genetic clustering are crucial for pinpointing these contributing factors. It is for this reason this study was conducted. This was a non-experimental, cross-sectional population-based molecular epidemiological study of TB in SC from 2005 to 2011. Its purpose was to estimate the proportion of TB that may be due to recently acquired infection and to determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates from TB patients in South Carolina from 2005-2011. The analysis sample included 627 confirmed pulmonary and/or pleural cases of TB, for which complete data on all covariates and a valid genotype were available. The results strongly suggested that about 50% of TB in South Carolina is recently transmitted. The study also revealed that being born in the United States and Black race were independently and significantly associated with being part of a TB genotype cluster. The key messages of this study were as follows: a substantial portion of TB in South Carolina is due to recent transmission, not reactivation or importation, and transmission of TB in South Carolina occurs in groups often defined by American birth and Black race. These important findings indicate that most TB in South Carolina is preventable and that enhanced TB control efforts should be explored. The implication for positive social change is that employing targeted contact investigation informed by these findings could lead to decreased disease transmission. Future studies should explore pilot programs that investigate alternatives to the traditional TB contact investigation.

The Molecular Epidemiology of Tuberculosis in South Carolina, 2005-2011: Estimates of Recent Transmission and Risk Factors for Genotype Clustering

by

Amy K. Roach

MSPH, University of South Carolina, 2003

BS, University of South Carolina, 2017

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2017

Dedication

This work is dedicated to the disease control staff of the South Carolina Department of Health and Environmental Control. You have taught and inspired me.

Acknowledgments

Thank you to my parents for instilling in me the work ethic, patience, and confidence to complete my journey. Thank you to my committee chair Dr. Ji Shen and committee member Dr. Amy Swango-Wilson for all of your expertise and guidance and to all School of Public Health faculty members at Walden University. Thank you to Dr. Kathryn Arden for your valuable subject matter expertise and mentorship.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Introduction

The global and domestic impact of tuberculosis (TB) disease is substantial. According to recent estimates, the current global incidence is nearly 9.6 million people (WHO, 2016). While global mortality has decreased in the last decades, it remains a staggering 1.5 million deaths in 2014 alone. This figure is unacceptable given that most of these deaths are preventable (WHO, 2016). The United States is considered a country with comparatively low incidence, at 3.2 cases per 100,000 persons. Despite this low case rate and the U .S. TB program, which is often regarded as the gold standard for control and prevention, between 500 and 600 people die every year from TB in the United States. Particularly concerning are the considerable racial and ethnic disparities extant in TB disease impact across the United States. In 2014, the rate of TB disease in Blacks was 5.8 cases per 100,000 persons, which is over 7 times higher than the rate of TB disease in White, non-Hispanics (0.8 cases per 100,000 persons). Similarly, in Hispanics/Latinos, the rate of TB disease was 5.3 cases per 100,000 persons–again, over 7 times higher than the rate of TB among White, non-Hispanics (CDC, 2016).

The Division of Tuberculosis Elimination (DTBE) at the CDC has, as its primary goal, the elimination of TB in the United States, where elimination is defined as ≤1 case/million persons (CDC, 2013). To this end, continuing research in U. S. populations affected by TB is imperative. Molecular studies of genotype clustering are important for identifying risk factors that contribute to the ongoing transmission of TB in the U.S.. Genotyping may facilitate quicker (a) confirmation of known contacts, (b)

detection of unknown contacts, or (c) revelation of transmission environments (Malakmadze, 2006,Yeo, 2006). This has been exemplified by clustering studies that have elucidated risk factors, which, in turn, serve TB prevention efforts by informing TB control programs where to target their limited resources (Miller, 2002; Suffys, 1997).. The goal of this research was to estimate the proportion of South Carolina TB may be due to recently acquired infection, and determine the risk factors associated with the genetic clustering of identical *M. tuberculosis*. This study has implications for social change including to further the TB elimination goal by elucidating risk factors for ongoing transmission in South Carolina, thereby informing the South Carolina TB Control Program and potentially other states' control efforts, particularly states with populations and public health resources that compare favorably to those of South Carolina.

This chapter will review the problem and purpose of the study, describe the research questions, assumptions, limitations, commonly used terms, and describe the conceptual framework of the study.

Background

Because traditional contact investigation often fails to reveal source cases, molecular epidemiology is an important adjunct to modern TB control programs (Cacho Calvo, 2005; Ellis, 2002). While several studies employing TB genotyping have been conducted, many have focused on urban areas such as Los Angeles, Vancouver, New York, and San Francisco (Barnes, 1997; Hernandez-Gardduno, 2002; Driver, 2006 & Cattamanchi, 2006, respectively) or European populations such as England, Spain, and

Italy (Love, 1998; Cacho Calvo, 2005; Francetti, 2010). The statewide studies in the U.S. have been conducted primarily in the northeast: Massachuttes, New York, and Maryland (Miller, 2002; Ellis, 2002; & Torgersen, 2006). The only study from the southeast U.S. was in Alabama on data from 17 years ago, 1994-2000 (Kempf, 2005). My study addresses this gap in the literature by characterizing the molecular epidmiology of TB in one state in the southeast United States, that of South Carolina. It provided a unique opportunity to examine both natively acquired and foreign-born tuberculosis, increasingly merging urban and rural environments, and conspicous, unaddressed racial disparities in TB disease (MMWR, 2006; MMWR, 2011). Because TB continues to elude elimination in the U.S., there is a need to identify contributing factors that may have implications for targeted control measures. Molecular studies of clustering are crucial for pinpointing these factors. While incident cases of TB have decreased in South Carolina over the previous 8 years, going from 261 cases reported in 2005 to 140 cases reported in 2011 (South Carolina DHEC, 2014), South Carolina still had the 15th highest case rate in the U.S. (2011) with 3 cases per 100,000 people (CDC, 2013). Of further concern is that the epidemiology of TB in South Carolina has long been characterized by extreme racial and ethnic disparities. From 2009-2014, 56.4% of South Carolina's TB cases were Black, 11.9% Hispanic or Latino, and 11.1% Asian, while Blacks, Hispanics/Latinos, and Asians comprised only 27.9, 5.1, and 1.3% of the state's population, respectively (CDC, 2016). Adding to the complexity of the race issue is the dramatic disparity in AIDS diagnoses in South Carolina, with 69% of new AIDS cases in 2013 occurring in Blacks (CDC, 2015). Because TB progresses from infection

to disease more often and more quickly in HIV patients, it is more difficult to treat and more likely to lead to death in persons co-infected with AIDS, illuminating and targeting any interacting racial, lifestyle, and economic disparities associated with these comorbidities should be a priority of public health. Approaches that could be initiated that simultaneously address TB and HIV co-infection should be investigated.

Another area of increasing concern in the last decade is the emergence and spread of multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB), both of which pose a global threat to health, economic development, and national security, undermining the significant progress made globally and domestically to accomplish TB elimination (United States Government Global Tuberculosis Strategy, 2015). South Carolina has been affected by this, having had nine cases of TB between 2005 and 2011 that were resistant to at least one TB treatment drug, only one of which was MDR-TB. Fortunately, South Carolina has had no cases of XDR-TB to date. My study population did not contain enough drug-resistant TB to examine drug resistance as a covariate. However, the emergence of MDR-TB worldwide and in the U.S. underscores the need for comprehensive genotyping in even low-prevalence communities such as South Carolina. Not because genotyping is necessary to distinguish drug resistance (other types of laboratory tests can determine drug resistance without the need for a full genotype) but because genotyping facilitates identification of chains of transmission. When and if cases of MDR-TB and XDR-TB occur, in order to target inventions it will be crucial for TB control programs to understand whether these cases were entirely imported or imported and then locally transmitted.

Statement of Problem

TB disease is a current, relevant, and immediate public health threat of global concern. Households with at least one income-producing member sick with TB but getting treated typically loses 3–4 months' work and about 2% of income; an untreated person with TB may lose as much as a full year of work. In the countries with the highest TB prevalence, which often happen to be the poorest countries, lost productivity due to TB may translate to 4–7% of the gross domestic product (GDP) (U.S. DOS, 2009). In the U.S., despite sustained efforts to eliminate TB and trends reflecting declining incidence in the last decade, more than 9,400 new cases were diagnosed in 2014 (CDC, 2015). Further, TB is not just a disease of adults, even in our wealthy modern society; children are still infected with 486 people less than 15 years old diagnosed in the U.S. in 2012 (CDC, 2013). TB infection in children is especially troubling because it represents new transmission (as opposed to latent disease), and thus indicates a conspicuous failure in U.S. TB control efforts. Moreover, as mentioned earlier, U.S. Blacks are over 7 times more likely to be diagnosed with TB than Whites are. Many reasons have been postulated for this disproportionately high rate of TB among Blacks compared to Whites, including SES, comorbidities, and genetics but it has yet to be entirely explained by any of those variables. SES may explain much of this disparity, but the relationship is complex, thus the solution will likely also be complex (Cantwell, 1998).

In order for public health to address TB's impact on children, minority populations, and income and productivity, the risk factors associated with transmission

have to be clearly defined. Cacho-Calvo and co-authors (2005) observed that public health contact tracers found epidemiological links in only 85 (37.4%) of the 231 cases belonging to a genotype cluster, indicating that traditional contact investigation misses many cases (). The findings of the Cacho-Calvo study underscore how important it is that genotype cluster investigations accompany traditional contact investigations (Cacho-Calvo, 2005). A Canadian study from 2006 had similar findings and revealed how essential TB genotyping in children can be. Although sputum specimens can be extremely difficult to obtain in children, Yeo et al. (2006) were able to obtain specimens and successfully genotype *M. tuberculosis* in 38 pediatric TB cases. From genotyping, investigators identified 14 possible source cases. In contrast, they were only able to identify one possible source case from traditional contact investigation alone (Yeo, 2006).

Studies that have examined predictors of TB genotype clustering have had some success in clarifying factors associated with TB transmission. Among confirmed TB cases, three characteristics were found by multiple studies to be associated with being part of a TB genotype cluster: history of incarceration, history of alcohol abuse, and history of illicit drug abuse (Barnes, 1997; Ellis, 2002; Kempf, 2005; Cattamanchi, 2006; Driver, 2006). Not coincidently, these are also characteristics associated with crowded, marginalized populations that may have limited access to healthcare. My research examined these factors and others to determine those that were significantly associated with being a part of an identical TB genotype cluster of two or more cases, in hopes of clarifying TB transmission dynamics in South Carolina.

Purpose of Study

The two goals of this research were to estimate the rate of recent TB transmission and to determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates from TB patients in South Carolina from 2005-2011 by using a multivariable logistic regression technique to model risk factors for the binary outcome of being part of a TB genotype cluster, yes or no.

Research Questions and Hypotheses

I developed the following research questions in order to determine the risk factors for TB genotype clustering, and to estimate the rate of recent transmission of TB in South Carolina from 2005-2011. The literature suggests that, based on population-based studies, TB isolates sharing identical genotype profiles (also known as clustered isolates) are likely from patients with recently acquired infection (Driver, 2006; Ellis, 2002). Therefore, the answer to Research Question 1 would approximate what proportion of TB in South Carolina may have been due to recent transmission.

Research Question 1:

a) Using the mycobacterial interspersed repetitive unit (MIRU) genotyping method, and spoligotyping, for cluster classification of tuberculosis cases in South Carolina, I estimated the proportion of TB cases that were genotyped clustered versus unique (hence forward referred to as "clustered versus singleton")?
b) I estimated the proportion of South Carolina TB cases that may be due to recently acquired infection. The following logic was applied:

I. *For genotype clusters of only two cases*: One case of the cluster was assumed to be a source case. One case of the cluster was assumed to be the recently infected case.

II. *For genotype clusters of two or more cases*: One case of the cluster was assumed to be the source case. All other matches in the cluster were assumed to be due to recent transmission.

Thus, *C - 1*, were counted as recently transmitted cases, where *C* was the number of identical isolates in the cluster. This method is based on a recent transmission index that has been used in prior studies, *RTIn-1 = (nc - c)/n*, in which *n* is the total number of the studied cases, *nc* is the total number of cases in a cluster (size 2 or greater) and *c* is the number of genotypes represented by at least two cases. Based on this index, patients in a cluster are considered recent transmission and non-cluster cases considered reactivation. The *n−1* approach denies the possibility of more than one source case. This index has also been referred to as the "*n-1 method*" (Reza Allahyar Torkaman, 2014; Ricks, 2009).

Research Question 2: I determined the risk factors of genotype clustering among incident South Carolina TB cases from 2005 to 2011 considering the following hypotheses:

H_0 There is no relationship between being US-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being US-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being US-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between having prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between having prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

The dependent variable of interest for this study was being part of an identical TB genotype cluster of two or more cases. The *primary* independent variables of interest for this study (or potential risk factors) were as follows:

being US-born, being younger, being male, being Black, history of homelessness, history of incarceration, history of alcohol abuse, history of illicit drug abuse, history of non-injection drug use, history of injection drug use, history of substance abuse (alcohol and all drug abuse combined), being HIV positive, and prior TB disease.

Based on previous research, it was hypothesized that having the following risk factors could independently predict being part of a TB genotype cluster: a history of incarceration, born in the United States, younger age, being male, being Black, being of younger age, history of homelessness, HIV positive, history of alcohol abuse, and a history of drug use (Barnes, 1997; Ellis, 2002; Kempf, 2005; Cattamanchi, 2006; Driver, 2006; Moonan, 2012).

Nature of Study

This was a nonexperimental, cross-sectional, population-based molecular epidemiological study of TB in South Carolina from 2005 to 2011. All data, including the TB genotype results, were secondary data. South Carolina DHEC staff originally collected all patient information and clinical specimens. The genotyping results came from TB case's isolates that were sent to a laboratory in Michigan that is under contract with the CDC to provide genotyping services to TB control programs in the US. Culture-confirmed cases of tuberculosis in South Carolina isolates' DNA patterns were analyzed using two polymerase-chain reaction (PCR) methods: MIRU and spoligotyping. Isolates considered genetically identical by both of these methods were be defined as clustered—the outcome (dependent) variable of interest for this study. Risk factors associated with being clustered were first examined individually in a

bivariate model (one independent variable and the dependent variable), then in a multivariable logistic model using a stepwise model building approach. The criteria for inclusion in the multivariable regression model was a *p-value* of < 0.10 and the criteria for inclusion in the final multivariable model was a *p-value* of < 0.05. All analyses were conducted using SAS 9.3.

Furthermore, clustered cases have traditionally been considered an indication of recent transmission (Driver, 2006, Ellis, 2002). This population-based, cross-sectional study examined the proportion of South Carolina tuberculosis that may be due to recent transmission, and discusses the implications of this for the South Carolina TB Control Program. The research population consisted of all incident cases of pulmonary and/or pleural, culture-confirmed TB in South Carolina from 2005 to 2011. Thorough contact investigations and directly observed therapy (DOT) were the standard of care for all cases of TB in South Carolina. During these processes, extensive case follow-up and data collection were performed by the staff of the South Carolina TB Control Program. All patients included in this study were reported to the CDC national case registry via the report of a verified case of tuberculosis ([RVCT], CDC DTBE, 2009). This form included 49 variables on patient demographics, laboratory test results, drug suceptiblities, clinical background, clinical outcomes, and risk behavior. (A copy of this form is included as Appendix A.) All clincial information, such as laboratory test results, drug suceptabilities, clinical background and outcomes, were reported by the TB nurse and verified by the TB consulting physician. Information on patient demographics and risk behavior was primarily self-reported by the patient. History of homelessness

was self-reported except in circumstances where the case was identified or treated in a homeless shelter. The TB nurse then marked this question affirmative.

Definition of Terms

The following section defines terms used throughout the study. Some terms are discussed further in the literature review.

1. *Acid-fast bacilli* (AFB): those microorganisms that when stained retain color even after they have been washed in an acid solution, and may be detected under a microscope in a stained smear.
2. *Active TB disease*: an illness, caused by bacteria called *M. tuberculosis*, in which bacteria are multiplying and attacking parts of the body, usually the lungs. A person with active TB disease is able to spread disease to others if the TB bacteria are active in the lungs or throat. The symptoms of active TB disease may include weakness, weight loss, fever, no appetite, chills, sweating at night, bad cough, pain in the chest, and coughing up blood.
3. *AIDS-defining condition*: is the list of diseases published by the CDC that are associated with AIDS, and used worldwide as a guideline for AIDS diagnosis. According to the CDC definition, a patient has AIDS if he or she is infected with HIV and has either:
    - CD4+ T-cell count below 200 cells/μL
    - a CD4+ T-cell percentage of total lymphocytes of less than 15%
    - or one of the defining illnesses.

4. *Concentric circle method*: a method of classifying and screening contacts in order of intensity of exposure and risk of being infected. Contacts with the most exposure or highest risk of infection are screened first.
5. *Congregate setting*: a setting in which a group of usually unrelated persons live in close physical proximity. These settings may include hospitals, long-term care facilities, assisted living facilities, prison, jails, or homeless shelters.
6. *Contact investigation* (may be referred to as contact tracing): a procedure for interviewing a person who has TB disease to determine who this person may have exposed to TB. Then those people who may have been exposed are tested for latent TB infection (LTBI) and TB disease.
7. *Contacts:* people exposed to someone with infectious TB disease, usually family members, roommates, close friends, and sometimes coworkers, classmates, 'drinking buddies', illicit drug use companions, and others.
8. *Directly observed therapy (DOT)*: a component of TB case management that helps to ensure that patients adhere to treatment, where the health care worker or another designated individual administers and watches the patient swallow every dose of the prescribed drugs.
9. *Drug-resistant TB*: TB caused by organisms that are unable to grow in the presence of a particular drug; TB that is resistant to at least one first-line anti-tuberculosis drug.

10. *Ethambutol (EMB):* a drug used to treat TB disease; may cause vision problems. Ethambutol should be used cautiously in children who are too young to be monitored for changes in their vision.
11. *Extrapulmonary tuberculosis:* TB disease where infection has spread outside the lungs, this may include the pleural space, the central nervous system, the genitourinary system, and the lymphatic system.
12. *First-line TB drugs:* the initial drugs used for treating TB disease. Include isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), and either ethambutol (EMB), or streptomycin (SM).
13. *Foreign-born person:* a person born outside the U.S. that currently resides in the United States.
14. *Foreign-born tuberculosis:* tuberculosis infection that was likely acquired outside the United States.
15. *HIV seropositivity*: testing positive for the presence of HIV antibodies in the blood indicating infection with Human Immunodeficiency Virus.
16. *Interferon-gamma release assay (IGRA)*: a type of blood test that measures a person's immune reactivity to *M. tuberculosis* by measuring release of IFN-γ. In the US, QuantiFERON®-TB Gold, QuantiFERON®-TB Gold In-Tube, and T-SPOT® TB are the currently available IGRAs brands.
17. *Isolate:* a sample from a specimen that was identified as a certain organism such as *M. tuberculosis complex*.

18. *Isoniazid (INH):* a drug that is used for treating LTBI and one of the drugs used to treat TB disease; although relatively safe, it may cause hepatitis and other severe adverse reaction in some patients.
19. *Latent tuberculosis infection (LTBI):* a state of persistent immune response to stimulation by *Mycobacterium tuberculosis* antigens without evidence of clinically manifested active TB. Persons with latent TB infection do not feel sick and do not have any symptoms. They are infected with *M. tuberculosis*, but do not have TB disease. The only sign of TB infection is a positive reaction to the tuberculin skin test or TB blood test. Persons with latent TB infection are not infectious and cannot spread TB infection to others.
20. *Mantoux tuberculin skin test (TST)*: a method of testing for TB infection sometimes referred to as a Purified Protein Delivery test (PPD), is where a needle and syringe are used to inject 0.1 ml of 5 tuberculin units of liquid tuberculin between the layers of the skin (intradermally), typically on the forearm; the reaction to this test, a palpable swollen induration, is measured 48 to 72 hours after the injection and is interpreted as positive or negative depending on the size of the reaction and the patient's risk factors for TB.
21. *Miliary TB*: TB infection from a histological or radiologic finding, rather than a site of disease. It appears on radiograph as many small, well-defined nodules that resemble millet seeds scattered throughout the lungs, hence the name "miliary." Usually a very serious type of infection.

22. *Multidrug-resistant TB (MDR TB)*: TB organism that is resistant to at least the drugs isoniazid and rifampin.
23. *Mycobacterial interspersed repetitive units (MIRU):* comprise short tandem repeat structures found at multiple loci throughout the *Mycobacterium tuberculosis* genome and are used for genotyping TB pathogens.
24. *Mycobacterium tuberculosis complex (MTBC)*: a genetically related group of Mycobacterium species that can cause tuberculosis in humans, the most common of which is *Mycobacterium tuberculosis*.
25. *Nucleic acid amplification (NAA):* a laboratory technique that amplifies (copies) DNA or RNA segments, to identify microorganisms in sputum specimens.
26. *Pleural effusion*: the abnormal accumulation of fluid in the space between the lungs and chest wall.
27. *Polymerase-chain reaction (PCR)*: a biochemical technology in molecular biology used to amplify a single copy or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.
28. *Polymorphism*: a natural variation in a gene, DNA sequence, or chromosome
29. *Pulmonary disease:* TB disease that occurs in the lungs which accounts for most (about 90%) active disease.
30. *Purified protein delivery (PPD) test:* a method of testing for TB infection. See Mantoux TST above.

31. *Recurrence*: a patient who has either a history of a (1) negative culture result while receiving anti-TB therapy, but then at some point after therapy is completed, either the culture result becomes positive for *M. tuberculosis* or the patient has clinical or radiologic deterioration that is consistent with active TB disease or (2) *negative smear and culture result at diagnosis* and while receiving anti-TB therapy, but then at some point after therapy is completed, either the patient has a culture result that is positive for *M. tuberculosis* or has clinical or radiologic deterioration that is consistent with active TB disease.
32. *Rifampin*: a drug used to treat TB disease; also used for LTBI treatment. Rifampin may have serious side effects such as hepatitis, turning body fluids orange, and drug interactions.
33. *Second-line TB drugs:* drugs used to treat TB that is resistant to first-line drugs such as capreomycin, kanamycin, ethionamide, cycloserine, ciprofloxacin, and amikacin.
34. *Singleton case:* a TB case with an isolate that has a unique DNA fingerprint (i.e. does not belong to a known genotype cluster).
35. *Smear*: a specimen that has been smeared upon a glass slide, stained, washed with acid solution, and then placed under the microscope for examination. It is used to detect acid-fast bacilli in a specimen.
36. *Spoligotyping*: spacer oligonucleotide typing, or spoligotyping, is a rapid, polymerase chain reaction (PCR)-based method for genotyping strains of the Mycobacterium tuberculosis complex (MTB).

37. *Sputum:* phlegm from deep in the lungs, collected in a sterile container for processing and examination.
38. *Tuberculosis genotype cluster*: two or more TB case-patients whose *Mycobacterium tuberculosis* isolates have matching spoligotyping and 12-locus mycobacterial interspersed repetitive unit–variable number tandem repeat (MIRU-VNTR).
39. *Tuberculosis genotyping*: a laboratory-based genetic analysis of the bacteria that cause TB disease and when combined with epidemiological data has sufficient discriminatory power to determine TB cases likely to be in the same chain of transmission or conclude that cases are not related.
40. *Extreme-drug resistant TB (XDR TB)*: the occurrence of TB in persons whose *M. tuberculosis* isolates are resistant to isoniazid and rifampin, plus resistant to any fluoroquinolone and at least one of three injectable second-line drugs.

Assumptions

This study was based on three assumptions. First, most of the social (sometimes referred to as "lifestyle") characteristics on the RVCT were self-reported by the case. The characteristics were assumed true for the purposes of this study. Some studies have indicated that self-reported drug and alcohol use are generally underreported due to the stigma associated with substance use disorders (Fendrich, 2004; Mensch, 1988; Midanik, 1988; Romelsjo, 1995). However, no studies on the reliability and validity of these specific questions from the RVCT have been conducted. But because TB health department nurses visit their patients 2 to 3 times per week for as long as 9 months (in order to

provide DOT), it cannot be ruled out that some of the lifestyle characteristics noted on the RVCT were suspected/witnessed and then verified by the TB nurse. By this, I mean that a patient may have denied drug abuse when questioned directly by the nurse, but over the course of several visits, the nurse may have suspected this behavior and corrected the information in the patient's medical chart as well as the accompanying RVCT (which was part of the patient's medical chart). The extent to which "correction" of self-reported data was or was not a systematic practice by TB nurses cannot be determined. Thus lifestyle factors, such as drug abuse, were assumed to be a sensitive question of self-report. Another lifestyle factor that would have been self-reported by the patient was alcohol abuse. However, the TB nurse would have taken special care to verify the accuracy of this answer because alcohol consumption (and especially abuse) is a contraindication to some TB treatment. Consequently, the accuracy of the alcohol abuse question in this data set was expected to be high.

The second assumption was that the reliability and validity of the RVCT were high; however, the literature review did not find any research that directly measured the reliability or validity of the RVCT. This is discussed more in Chapter 3, Methodology.

Thirdly, that only active TB cases were counted in TB prevalence. People with latent tuberculosis infection (LTBI) are not symptomatic, not contagious, and not counted in TB disease prevalence rates. When they reactivate, they become active TB cases and thus would be more likely to be detected by the South Carolina TB control program, as they would manifest clinically diagnosable symptoms. Asymptomatic contacts with positive skin tests that were treated to prevent progression to TB disease

did not meet the CDC case definition for an incident or prevalent TB case. This was a valid assumption and is true for most TB prevalence studies.

Limitations

This section will discuss the limitations of this study. The first limitation of this study was that it is limited to those confirmed TB cases for which a valid TB genotype could be obtained. There are two reasons a confirmed case may not have a valid TB genotype. The first reason is when a TB case is confirmed without the benefit of a TB positive culture (recall that a genotype cannot be performed without a TB cultured isolate to perform it on). This is true for all cases that are exclusively extrapulmonary (TB cases that have only extrapulmonary disease rather than both pulmonary and extrapulmonary disease) because these cases are typically confirmed via some other testing methodology, such as a PCR. This is true with the exception of extrapulmonary cases that are of the pleural space of the lungs, which are usually confirmed, similar to pulmonary cases, using sputum specimens. In fact, for all practical purposes are epidemiologically similar to pulmonary TB. Thus, extrapulmonary cases of the pleural space were considered eligible cases for genotype in the research sample as well as contagious cases that have contributed to the chain of transmission in the community. Confirmed cases that may also not have a TB culture are those pulmonary/pleural cases that are confirmed by some combination of symptoms and testing other than a positive sputum culture. There are a few reasons why a person with true TB pulmonary/pleural disease may not have had an *M. tuberculosis* organism collected and/or isolated. These include, but are not limited to, (a) the patient being lost to follow-up, thus no specimen was collected; (b) the patient received partial

antibiotic treatment and thus had no viable organism in the specimen; (c) poor specimen collection in that the specimen that is collected is not sputum from the lung but was shallow esophageal mucous or saliva where little to no organism was present; and (d) poor specimen shipping or transfer procedures resulting in viable organism expiring before it could be cultured and identified; and (e) the patient being between the ages of 5 and 12 years when a sputum specimen is not only difficult to obtain but is also difficult to isolate, thus alternative testing is often used to confirm these cases.

The second reason a confirmed case may not have a valid TB genotype was because all TB isolates are not being genotyped. Ideally, all South Carolina TB diagnosing laboratorians would send all their TB isolates to the National TB genotyping lab in Michigan for genotyping. The goal for all states participating in the National CDC Genotyping Surveillance program (of which South Carolina is one) in 2014 was 90% participation and the goal for 2020 is 100% participation. However, South Carolina has been falling well short of this goal. From 2005 to 2011, there were 1,346 confirmed cases of TB in South Carolina. Of those, 1,080 were pulmonary/pleural only or both pulmonary/pleural and extrapulmonary, thus 1,080 were eligible to be diagnosed by a sputum culture. A valid genotype was available for only 685 (63%) of those cases.

In summary, the missing 37% of confirmed TB cases was a combination of four possibilities: Either

- A culture was not obtained (as mentioned above, this happened frequently in children, and occasionally patients were lost to follow-up).

- *M. tuberculosis* was not isolated (again, this happens more frequently in children and may happen if a patient has received partial antibiotic treatment).
- The isolate was not sent for genotyping (the South Carolina lab did not follow protocol).

or

- The isolate was not able to be successfully genotyped.

Unfortunately, in my study sample I had no way of knowing what reason or combination of reasons a confirmed TB case patient may not have had a successful isolate and/or genotype. In the dataset I was provided by South Carolina DHEC, I had covariate information for all the TB cases for which I also had a valid genotype. I had no way of knowing in what ways those 420 (37%) incident TB cases without a valid genotype differed from the 685 incident TB cases for which I did have a valid genotype. There was no reason to believe these 685 cases were randomly selected for genotyping. This may have introduced some bias into the estimate of recently acquired infection and in the determination of risk factors for genotype clustering. To mitigate this limitation I used publically available descriptive data on the TB incident case population of South Carolina from 2005 to 2011 to compare with the fully genotyped sample on the covariates of interest. These demographic variables were available through an intranet query site called the CDC's Online Tuberculosis Information System (OTIS) that provides data in 5-year summary totals by state (CDC, 2016). The information from this site allowed me to determine how alike or different my sample was from the entire sampling frame, and to qualify my conclusions appropriately.

In addition, this limitation also represented an opportunity. The opportunity to emphasize the importance of comprehensive TB case genotyping, to characterize the molecular epidemiology of TB in any jurisdiction or geography. This limitation and its implications are further discussed in the recommendations section of Chapter 5.

A second limitation of this study was that any contribution that unconfirmed symptomatic cases could play in the TB chain of transmission, and how this affected my estimate of recent transmission, was unknown. Because some symptomatic undiagnosed cases (those who never seek medical care and are under the health department's radar) will contribute to the chain of transmission, how these cases may or may not differ with the study population regarding the independent variables of interest cannot be explored. Therefore, any bias this introduced cannot be examined. However, even the most robust TB control programs will miss cases. Again, how these cases differ from the cases that are detected by public health surveillance systems are not entirely clear. These cases may be more likely to be transient, underserved by the healthcare community, possibly undocumented immigrants, or cases of subclinical manifestation. This is a limitation for most TB genotyping prediction studies.

This third limitation of this study was that it was limited in place and time, in that I was only able to examine genotypes of cases diagnosed within South Carolina between 2005 and 2011. Cases diagnosed outside of this time period (either before or after) within South Carolina; cases diagnosed within this time period but outside of South Carolina; and cases diagnosed outside this time period (either before or after) and outside of South Carolina, were not included in this study. Therefore, TB chains of

transmission beyond the geographic and time boundaries set in this study were not investigated. Certainly, this is a common limitation of all molecular epidemiological studies, because researchers will always be limited by the data they collect or, as in the case of this study, that which is secondarily available to them. However, because TB transmission recognizes no time or state boundaries, this has implications for the conclusions that may be drawn from this study's findings. They are explored in Chapter 5, Conclusions, of this research.

The fourth limitation of this study was that while it was a population-based study, it represented only individuals living in South Carolina. The fact that the study was confined to South Carolina has implications for the generalizability of the study results. While South Carolina has a growing urban population, a few modest-sized airports with some international flights (such as to Canada and Mexico), and some public transportation, it is somewhat behind many other states with regard to these factors, especially those in the U.S. northeast. Thus, South Carolina may not be comparable to U.S. urban northeast in terms of variety of TB, the presence of drug-resistance TB, and conditions of urban overcrowding that may increase risk of transmission for TB. Using public transportation and living among concentrated urban populations have been shown to be risk factors for TB infection (Weis, 2002; Friske, 2011; Kirenga, 2015). Further, South Carolina has few direct commercial flights from some of the highest TB incidence regions, such as Africa and Indonesia. For these reasons, this study may not be entirely comparable to U.S. states with (a) more concentrated urban populations, (b) expansive, highly used public transportation systems, and (c) large international

airports, because these three factors are likely to influence the amount and variation of TB genotypes and clustering.

A fifth limitation of this study was that, for the purposes of answering Research Question 1(b), what proportion of TB in South Carolina may be due to recent transmission?, epidemiological information (dates of onset, diagnoses, etc.) was not be reviewed. This determination was a simple mathematical calculation that provided a reasonable indication of the proportion of South Carolina TB cases that could be due to recent transmission. It was beyond the scope of this study to carry out an in-depth review of source medical records to determine the true source for comparison to the mathematically calculated determination. First, because the time it takes latent TB infection (LTBI) to manifest as disease varies widely because it is often a result of both known and unknown host factors. The dates of onset and diagnosis obtained may not be helpful or reliable in determing source or index cases for clusters or outbreaks. Second, most TB population-based molecular epidemiological studies do not use medial record review to determine the proportion of recent transmission, but do it mathematically in just the way I have outlined in Research Question 1(b). This was an acceptible limitation because the methods used in this study were consistent with previous research and my research is directly comparable to other studies of this nature.

Significance of the Study

One of five priority actions that the World Health Organization (WHO) has outlined as necessary to accelerate progress towards the 2020 Global Tuberculosis targets is to *reach the missed cases*. From a global perspective, about 3 million people who

developed TB in 2012 were missed by national notification systems (WHO, 2013). As part of the positive social change implications of this research, it is my hope that cases that would have been missed when investigated with traditional contact investigation alone will be discovered when employing targeted contact investigation informed by the findings of this study. Genotyping may facilitate quicker confirmation of known contacts, detection of unknown contacts, or revelation of transmission environments (Malakmadze, 2006, Yeo, 2006). If a jurisdiction participates in the U.S. national TB genotyping program, consistently sends in all its confirmed case specimens for genotyping, and then checks those results frequently, they may find linked cases that their standard contact investigations did not reveal. In addition, TB contact tracers may confirm suspected epidemiological links they were already investigating. This approach may help a jurisdiction make decisions about how aggressive and widespread their contact investigations should or should not be, where to focus limited resources, and whether there are transmission environments that traditional contact investigations are missing. This has been exemplified by clustering studies that have elucidated risk factors that serve TB prevention efforts by informing TB control programs where to target limited resources (Miller, 2002; Suffys, 1997).

Furthermore, this study has two long-term implications. First, these findings may lead to better TB control efforts and thus interrupt the chain of transmission yield fewer TB infections. This would bring the U.S. one-step closer to TB elimination. Second, targeted contact-investigation informed by TB genotype results may yield an overall cost-savings to the South Carolina TB Control Program and similar programs.

Summary

TB is a current and immediate public health threat of worldwide importance. The DTBE at the CDC has, as its long-term goal, the elimination of TB in the US. To this end, my research examined factors that significantly contribute to being a part of an identical TB genotype cluster in South Carolina between the years 2005 and 2011. As part of the positive social change implications of this research it is my hope that cases that would have been missed when investigated with traditional contact investigation alone, will be discovered when employing targeted contact investigation informed by the findings of this study. Furthermore, the long-term implications of this study may lead to better TB control efforts with targeted contact investigation that results in an overall cost-savings to the South Carolina TB Control Program and similar programs.

Chapter 2 will provide an in-depth discussion of the research on U.S. and European populations with low incidence of TB that is similar to my study population. Chapter 2 will also provide a framework for this research and describe the gap in the literature that my research fills.

Chapter 2: Literature Review

Introduction

This research examined (a) what proportion of South Carolina TB cases from 2005 to 2011 were due to recent transmission and (b) the risk factors among incident cases that were significantly associated with being part of an identical TB genotype cluster of two or more cases. The goal of providing additional clarity to South Carolina TB transmission dynamics, and useful information to the South Carolina TB Control Program. To frame this research and identify gaps in the literature, I reviewed applicable studies on TB genotyping and the predictors for clustering in relevant populations. While there have been population-based studies examining predictors for clustering in the U.S. and other developed countries as well as the developing world, I found a paucity of studies in the southeast U.S.. This reflected an important gap in the literature. The southeast U.S. remains an area of high TB incidence, particularly among Black people. In the southern states, TB is almost exclusively transmitted in relatively insular networks defined by race, ethnicity, and SES (Moonan, 2012; Kempf, 2005). The reasons for this are not clear, but in general, southern states provide less Medicaid funding per person; there is less funding for infectious disease control and intervention; and there is less direct state funding to public health departments (Salinsky, 2010; Reif, 2012; DHHS, 2013). Further, remnant and non-traditional segregation may also play a role in this racial disparity for what is a highly infectious disease transmitted person-to-person. These concerning questions indicated a need for continued research. Certainly, the number of individuals suffering from, or at risk of, TB in the developing world—as well in the crowded urban areas of the

developed world of the US, Canada, and Europe—has prompted some important TB genotyping studies in these geographic regions (Verner, 2004; Moonan, 2012; Hernandez-Garduno 2002; Kamper-Jørgensen, 2012,). However, this review was limited to those studies that examined predictors for clustering in populations that were similar to my study population, such as those in other industrialized countries that could be expected to have similar TB incidence as the US.

The review begins with an overview of the search strategy. I then discuss the history of tuberculosis. The first part of the review discusses the conceptual framework behind tuberculosis genotyping. I then discuss the rationale for in the chosen statistical analysis methods. Following that, I review the literature related to each of the covariates I examined in my study. Finally, I discuss a review table of the various rates of clustering and recent tuberculosis transmission observed in prior population-based genotype clustering studies.

Literature Search Strategy

In this literature search, the following databases were used: CINAHL, Medline, Health Sciences, ProQuest Health and Medical Complete, Ovid Nursing Journals, and PubMed. I limited the searches to peer-reviewed journals published within the last 12 years. (Seminal research older than 12 years was also reviewed to provide historical context.) In all searches, I used the following keywords or phrases: *tuberculosis genotype clusters, tuberculosis genotyping, tuberculosis clusters, molecular surveillance of tuberculosis, molecular epidemiology of tuberculosis,* and *DNA fingerprinting of tuberculosis*.

Tuberculosis Background

The causative agent of most human TB disease is an aerobic bacterium that was discovered by Robert Koch in 1892, and later named *Mycobacterium tuberculosis*. The *M. tuberculosis* complex (MTBC) includes four other TB-causing mycobacteria: *M. bovis, M. africanum, M. canetti,* and *M. microti*. *M. africanum* is not widespread, but it is a significant cause of tuberculosis in parts of Africa. *M. bovis* was once a common cause of tuberculosis, but the introduction of pasteurized milk has largely eliminated this as a public health problem in developed countries. *M. canetti* is rare and seems to be limited to the Horn of Africa, although a few cases have been seen in African emigrants. *M. microti* is also rare and is mostly seen in immunocompromised people (CDC, 2011). Historical research indicates that by the time the causative agent was discovered every seventh person in the world was likely already infected (Santic, 2013). The source of infection is typically a person symptomatic with pulmonary, laryngeal or bronchial TB that then transmits by infectious droplets to their close contacts. The most infectious person is someone who in 1 milliliter (ml) of sputum will excrete around 10,000 TB germs, and this sputum will be TB positive on microscopic tests. Because TB is acquired through airborne transmission of droplet nuclei risk of infection has been shown to increase with nuclei concentration in droplets and with time of exposure to these nuclei (Bass, 1990). Once an individual has been infected, s/he remains infected for a long time, possibly progressing to active disease, sometimes years after the initial infection. In healthy people, about 10% of infected cases will progress to active disease. Conditions such as immunosuppression with HIV, physical and emotional stress, and very young age all substantially increase risk for

developing active disease after primary infection, the risk being highest among young children. Development of TB disease may be due to reactivation of latent disease or to reinfection. Effective treatment for TB was not developed until 1944 when streptomycin was first used. This was later followed by isoniazid and rifamycin regimens, which significantly increased the TB cure rate to as much as 95% (Suffys, 1997).

Most TB infection is concentrated in the lungs; however, extrapulmonary tuberculosis is TB disease where infection has spread outside the lungs. This may include the pleural space, the central nervous system, the genitourinary system, the lymphatic system or the skeletal system. It may be more difficult to diagnose and treat. It is more common in HIV-infected patients because it is related to the failure of the immune response to contain *M. tuberculosis*, thereby enabling haematogenous dissemination and subsequent involvement of single or multiple nonpulmonary sites (Lee, 2015). The organism proliferates and disseminates throughout the body ("miliary" tuberculosis). Cough may not be a typical symptom as the initial pulmonary infection may have passed by this point. Symptoms of extrapulmonary TB are vague and include fever, weight loss, night sweats, anorexia, and weakness. Extrapulmonary TB is much more common in the developing world where no treatment, inadequate treatment, and treatment failure are more common than in the U.S. (Lee, 2015).

Conceptual Framework

Because prompt identification of TB patients through symptom screening and testing, along with evaluation of contacts, can be difficult in hard-to-reach populations, thorough identification of TB contacts is more important than ever to achieving the

national goal of TB elimination (Dade County Cluster, MMWR 2012). Genotyping may facilitate quicker confirmation of known contacts, detection of unknown contacts, or revelation of transmission environments (Malakmadze, 2006; Yeo, 2006). If a jurisdiction participates in the U.S. national TB program, consistently sends all their confirmed case specimens in for genotyping, and checks those results frequently they may find linked cases their standard contact investigations did not reveal. In addition, TB contact tracers may confirm suspected epidemiological links they were already investigating. Importantly, this approach may help a jurisdiction make decisions about how aggressive and widespread their contact investigations should or should not be, where to focus limited resources, and if there are transmission environments their traditional contact investigations are missing. This has been exemplified by clustering studies that have elucidated risk factors, which serve TB prevention efforts by informing TB control programs where to target limited resources (Miller, 2002; Suffys, 1997). Because the U.S. national TB genotype coverage (i.e. the proportion of confirmed cases that are successfully genotyped) has increased from 51.2% in 2004 to 88.2% in 2010 (MMWR, 36, 2012), this has greatly expanded genotyping's value in characterizing populations at high risk for TB transmission and outbreaks.

The primary concept applied in this study is commonly referred to as tuberculosis genotyping. Most simply defined, TB genotyping is the laboratory-based genetic analysis of the bacteria that cause TB disease, and when combined with epidemiological data has sufficient discriminatory power to help find TB cases likely to be in the same chain of transmission, or determine that cases are not in the same chain

of transmission. There are three primary methods for genotyping TB isolates: IS-6110-based genotyping also known as restriction fragment length polymorphism (RFLP) technique, MIRU-based genotyping, and spacer oligonucleotide typing also known as spoligotyping-based genotyping. For this review, I will focus on MIRU-based genotyping and Spoligotyping because those are the methods that were used in this study. However, in this section, I will provide a brief discussion of RFLP technique as it was used in some of the studies I will reference and it was a groundbreaking method for genotyping.

One of the first methods for TB genotyping developed in the early 1990s was RFLP technique, and it has been used extensively for TB genotyping and epidemiological studies since that time. There is a substantial amount of research indicating the accuracy of RFLP in determining TB genotype matches (Samper, 1998; Barnes, 2003; Ellis, 2002; Love, 2008). It is the standard approach for the analysis of the distribution of the insertion sequence IS6110 in different strains. Its basis is that isolates from patients infected with epidemiologically unrelated strains of tuberculosis have different RFLP patterns, or different distribution sequences of IS6110, whereas those from patients with epidemiologically linked strains generally have identical RFLP patterns. While it is a highly discriminatory method, it is also complex and time consuming, as it requires sub-culturing isolates for several weeks to obtain sufficient DNA for typing. Currently, large databases of IS6110-based genotypes are available for TB control programs and researchers to review and use to compare strains. Because strains with fewer than six IS6110 insertion sites have a limited degree of

polymorphism, a supplementary method of genotyping may be required in this circumstance, such as spoligotyping.

Spoligotyping was developed somewhat concurrently with RFLP and has been used successfully in TB research for determining DNA patterns of *M. tuberculosis* isolates (Barnes, 2003; Kempf, 2005). The direct-repeat locus in *M. tuberculosis* contains 10 to 50 copies of a 36-bp direct repeat, which are separated from one another by spacers that have different sequences. However, the spacer sequences between any two specific direct repeats are conserved among strains. Because strains differ in terms of the presence or absence of specific spacers, the pattern of spacers in a strain can be used for genotyping. Spoligotyping has two advantages over RFLP. First, because only small amounts of DNA are required, it can be performed on clinical samples or on strains of *M. tuberculosis* shortly after their inoculation into liquid culture and thus it has a faster turn-around time. Second, the results of spoligotyping are expressed as positive or negative for each spacer thus they can be reported in a digital format (like the results of MIRU analysis noted below) facilitating the creation of large web-based databases. The primary limitation of spoligotyping is that it is less discriminatory than either RFLP or MIRU. However, when used in conjunction with MIRU, as it was in my study population, it has excellent matching power (Barnes, 2003).

A third method is MIRU-based genotyping. The genome of *M. tuberculosis* contains many MIRUs, some having identical repeat units and others having repeats that differ slightly in sequence and length. MIRU genotyping describes the number and size of the repeats using a polymerase-chain-reaction (PCR) assay, followed by gel

electrophoresis. The discriminatory power of MIRU genotyping is nearly as good as that of RFLP but unlike RFLP, MIRU analysis can be automated and many strains typed simultaneously, yielding results that can be digitally catalogued in a web-based database. MIRU is technically simpler than RFLP and can be applied directly to *M. tuberculosis* cultures without DNA purification first, thus resulting in a quicker turn-around. Because of the quicker turn-around of both the spoliogotyping and MIRU methods, these are better for use in ongoing contact investigations than compared with the RFLP method.

There is a substantial amount of research indicating the applicability of TB genotyping in TB epidemiological investigations, especially in low to moderate incident countries (Samper, 1998; Barnes, 2003; Ellis, 2002; Kempf, 2005,; Malakmadze, 2005; Love, 2008). For example, in a U.S. study Malakmadze and co-investigators revealed three clusters of 19 patients by matching patient isolates with all three methods: RFLP, MIRU, and spoligotyping. Then researchers retrospectively performed medical record reviews and patient interviews, which revealed that most of these clustered patients had no obvious epidemiologic links, but the medical records did point to several previously unrecognized locations of possible TB transmission. These unrecognized locations of transmission were a single-room occupancy hotel, two homeless shelters, one bar, and two crack houses. This study perfectly illustrates that transmission of TB among high-risk groups may go undetected for years when relying on patient recall alone. This is because it is very difficult to obtain a complete contact list from persons with often numerous and frequent transient living and socializing environments coupled with

alcohol and drug abuse behaviors. In addition, this investigation particularly highlights the value of using multiple methods for TB genotyping to confirm genotype matches, and reveal previously unrecognized locations of transmission that may be targeted for specific TB interventions (Malakmadze, 2005).

Statistical Analysis in the Literature

After a thorough review of population-based studies that have examined the association of various risk factors for clustering, I found that most studies consistently employ logistic regression to model this relationship. For the remainder of this chapter, all correlational studies reviewed should be assumed to have used this technique unless otherwise noted. Further, logistic regression was used as my primary data analysis technique for Research Questions 2. Logistic regression was an appropriate choice because it is a powerful and reliable statistical too and has been used for many decades in health science research (Kleinbaum, 1988; Hosmer, 1989). The nature of these research questions and the extensive use of logistic regression in the current literature to determine the impact of independent variables have on the dependent variable indicate it was the best option. Parsimonious model building via backward elimination of insignificant covariates was applied when calculating adjusted odd ratios (aOR). Odds ratios >1.0 with an associated *p-value* <0.05 were interpreted as a risk factor for clustering. Likewise odds ratios <1.0 with an associated *p-value* <0.05 were interpreted as protective for clustering. Further details regarding the statistical methodology used in my study are provided in Chapter 3.

Literature Review Related to Key Variables

Numerous studies have evaluated various demographic, social, and clinical

factors on risk for clustering versus singleton, where singleton is defined as a TB case that does not belong to a known genotype cluster (Love, 2009; Barnes, 1997; Cacho-Calvo, 2005; Cattamanche, 2006; Fok, 2008; Kempf, 2005; Chan-Yeung, 2006; Driver, 2006; Ellis, 2002). The body of literature that precedes my research used both first and second-hand data collection from a variety of sources including, but not limited to: the RVCT exclusively, the RVCT supplemented by retrospective medical record review, first-hand comprehensive medical record review, first-person patient interviews and medical record review, interviewer-administered questionnaires, or patient self-completed questionnaires. For this reason, the studies discussed in this literature review represent a wide range of covariates that have been collected and examined in relation to clustering. As with any research, my research was limited to evaluating only those factors (i.e. covariates) for which my secondary data set contains complete and reliable information. To those ends, the scope of this literature review and the list of covariates below is limited to only those that examined in my research.

Two seminal studies worth highlighting are that of Fok's 2008 meta-analysis on 36 population-based TB genotyping studies and Moonan's 2012 US-wide comprehensive genotyping and geospatial scanning estimate study. Both are referred to multiple times in this chapter. Fok's study is important because of its comprehensive meta-analysis of prior research (Fok, 2008). Moonan's study is important because it examined all TB cases in the U.S. that had a genotyping result from 2005 to 2009, and employed a geospatial scanning strategy to determine if matching genotype cases were likely to be a result of recent transmission (Moonan, 2012.)

Fok's meta-analysis (which does not include any studies after 2007) found that in countries with low TB incidence the characteristics of local birth, male sex, minority race, substance abuse (alcohol abuse and injection drug use), and homelessness were associated with TB clustering and recent TB transmission (a finding consistent with other published studies). Similarly, in the US, Moonan found that the characteristics of persons U.S. born, being male, members of a minority race or ethnic group, persons that abuse substances and the homeless are at higher risk for TB clustering and recent transmission, again a finding that is consistent with other U.S. studies.

Below, I have listed each covariate I examined in my study and a brief description of what the body of literature has found in relation to clustering. Each covariate is listed by the term I will use to refer to it for the remainder of this dissertation. In parentheses beside the covariate is the category/alternative of that covariate that previous research has usually observed (but not always) to be positively predictive for clustering. Generally, studies have grouped covariates into the broad categories of (1) demographic (2) social and (3) clinical. There may be some overlap between categories, but I have taken care to group each of my covariates similarly.

Demographic

Country of Birth (Not Being Foreign-Born). For studies conducted in the U.S., U.S. natives are more likely to be part of genotype clusters than those of foreign birth. This has also been observed in European studies, such that persons native to the country of study origin are more likely to be part of a genotype cluster than those of foreign

birth. For example, researchers conducting a study in the Netherlands found native-born Danes at higher risk for clustering compared to those of foreign birth, and researchers conducting a study in Italy find those native to Italy at higher risk for clustering than immigrants (van Soolingen, 1999 & Moro, 2002, respectively).

A Massachusetts study by Miller and colleagues found an adjusted odds ratio (aOR) of 2.29 (95% C.I. 1.69-3.12) for the association of clustering among U.S. born when compared to foreign-born persons (2002). Additionally, Moonan's a population-based study (including all genotyped cases of TB reported to the U.S. National Tuberculosis Surveillance System (NTSS) by the 50 states and the District of Columbia) observed an aOR of 2.4 (99% C. I. 2.1-2.7) between clustering and U.S. birth (2012). In Denmark, Kamper-Jorgensen and co-authors found that TB cases in large clusters (≥ 0) were 7.7 (3.6 -16.4) times more likely to be Danish born than non-clustered TB cases (2012). Similarly, in a study from the Netherlands, van Soolingen and colleagues observed that not being Dutch-born was protective for clustering in both Mediterranean and African-born residents with aORs of 0.7 (95% C.I. 0.6-0.8) and 0.7 (0.6-0.9), respectively. They also found that longer periods of residence in the Netherlands (> 2 years) was a risk factor for clustering with an aOR of 1.4 (95% C.I. 1.1-1.8) (van Soolingan, 1999). In Italy (among non-AIDS patients), the aOR was 1.44 (95% C.I. 1.08-1.92) for the association of native-born Italian nationality and clustering (Moro, 2002).

The reasons for this association are somewhat intuitive. First, persons born in countries with high TB incidence are more likely to have acquired TB in their country of origin prior to arriving to U.S. or Europe, as Western countries have comparatively low

incidence. Thus, foreign-born persons may be more likely to have a singleton genotype if their TB is diagnosed early after arrival. Second, native-born persons would likely have had more time within their native country to have acquired and spread TB to their contacts, than recently arrived immigrants. Essentially, foreign-born persons are more likely to have acquired their TB natively within a genotype cluster in their country of origin and may or may not propagate that cluster once they immigrate.

Age (younger). Most studies have found that younger age is a strong, independent predictor of clustering. How age is categorized and the reference group varies by study, but generally studies have observed that younger cases are more likely to be clustered while older cases are less likely to be clustered (Kamper-Jorgensen, 2012; Zolnir-Dovc, 2003; Barnes, 1997; Talarico, 2012; Moonan, 2012). In the Moonan population-based U.S. study (that included all genotyped cases in the 50 states and DC from 2005-2009), when investigators used 25-44 years as the reference age group they found that age 0-4 years were positively associated with clustering (aOR = 3.1; 99% CI. 1.4-6.8) and >= 65 years was protective for clustering (aOR = 0.5; 99% C.I. 0.4-0.6) (Moonan, 2012). Further, in an early study from central Los Angeles, Barnes and colleagues observed a similar finding with an aOR of 4.1 (99% C.I. 1.1-15.1) associated with younger age (1997). Another study from Arkansas found that among TB cases <65 years old 56.4 % were clustered versus those >= 65 years old 43.6% non-clustered with *p-value* <0.0001 (Talarico, 2012). European studies have observed comparable results. In Slovenia, investigators observed that clustering rate decreased as age increased from 46.4% (age group under 35 years) to 19.5% (age group above 65 years) yielding an aOR of 0.42

(95% C.I. 0.74-2.21) (Zolnir-Dovc, 2003). Maguire and colleagues observed a marked dose response relationship with age and clustering in Londoners from 1995 to 1997. When using 60 years of age and older as the reference group, authors observed those 0-19 years at the highest risk of clustering with an aOR of 2.65 (95% C.I. 1.59-4.44), those 20 to 34 at the second highest risk with an aOR of 1.51 (95% C.I. 1.02-2.22) and those 35-59 at the third highest risk with aOR 1.43 (95% C.I. 0.97-2.11).

The reason for the association of younger age and clustering may be that older cases are more likely the result of reactivation of latent infection. In addition, it is important to keep in mind that genotype studies must examine clustering over a limited period. Thus, while older persons might be part of clusters, the ability of studies to detect clusters in excess of their review period would be limited, and certainly, the use of the genotyping in general to detect clusters originating 20 to 40 years ago would not be possible as the technology did not exist. In summary, older people have had more cumulative time to have been infected with TB and for this reason their disease is more likely to be the result of reactivation than their younger counterparts; meaning the odds of infection having been acquired before study initiation, outbreak detection, or cluster propagation are greater in older persons than in that of younger persons.

Sex (male). For all studies where a sex association was observed, male sex was independently associated with being part of a TB genotype cluster. While not all studies have observed an association with sex and clustering (Franzetti, 2009; Driver, 2006; Maguire, 2003; Cacho-Calvo, 2005). However, no studies (to date) have observed a significant positive association between clustering and being female. The US-wide study

by Moonan and colleagues (referred to earlier) found that men were 1.2 (95% C.I. 1.1-1.3) times more likely to be clustered than women. In an ethnically diverse region of southern California, Rodwell and co-authors observed an aOR of 1.57 (95% C.I. 1.17-2.10) for men and clustering. Similarly, in Arkansas, Talarico and co-investigators found that 48.2% of men were clustered versus only 28.0% of women with a *p-value* of 0.058 (2012). In Kamper-Jorgensen's study from Denmark investigators observed that, men were 2.5 times more likely to be clustered than women (2012). Results are similar in direction and magnitude in the Ellis U.S. study of 7 sentinel surveillance sites, the Driver study of New York City, and the van Soolingen study in the Netherlands (2002, 2006, and 1999, respectively).

The reasons males are at higher risk for clustering are not entirely understood but probably work in concert with many of the same reasons men are at greater risk for TB in general, including increased rate and effect of alcoholism, increased rate of incarceration, more homelessness, more use of homeless shelters, greater delays in seeking treatment, and possibly more social/congregate alcohol and drug abuse behavior (Oeltmann, 2009).While most studies do control for some of these factors, no one study or combination of studies could be expected to entirely control for these complex social and clinical factors.

Race (Black). U.S. studies have consistently found that black race is a strong independent predictor for being part of a TB genotype cluster. Most U.S. studies use non-Hispanic white as a reference group when examining this association. In Moonan and co-authors' study that included all genotyped cases in the U.S. from 2005 to 2009,

investigators observed the following associations with clustering: an aOR of 2.4 (99% C.I 2.2–2.7) for black, non-Hispanic; an aOR of 1.7 (99% CI 1.5–2.0) for Hispanic/Latino; an aOR of 2.6 (99% C.I. 1.5–4.4) for native Hawaiian/Pacific Islander; and an aOR of 1.5 (99% C.I. 1.3-1.8), when using non-Hispanic white as the reference group. Talarico and co-authors observed similar results in an Arkansas cohort from 1996-2003 finding an aOR of 2.07 (95% C.I. 1.52-2.82) for non-Hispanic blacks and clustering when compared to non-Hispanic whites (2011). Likewise results were similar in Texas where Serpa observed TB cases clustered in 82% and 77% of blacks and whites, respectively ($p = 0.46$) and additionally that cluster size was significantly larger in U.S. born blacks than whites ($p < 0.001$) (2009).

Ethnicity (Hispanic/Latino). The association between being Hispanic/Latino and clustering is not as clear as that of black race. A very early study from San Francisco found in patients under 60 years of age being Hispanic to be a greater predictor (OR = 3.3, *p-value* = 0.02) for clustering than being black (OR = 2.3, *p-value* = 0.02) (Small, 1994). While one important comprehensive U.S. study observed Hispanic/Latino a statistically significant positive predictor for clustering, although less in magnitude than black race, that of Moonan noted above. A few studies have suggested a protective relationship with being Hispanic/Latino and clustering (Ellis, 2002; Weis, 2002).

Social

Homelessness (current, or within prior 12 months). Some studies have found that being homeless in the year prior to TB diagnosis is a strong independent predictor for being part of a TB genotype cluster (Moonan, 2012; Zolnir-Dovc, 2003; Ellis, 2002; Love 2009;

Barnes, 1997; Driver, 2006; Rodwell, 2012). While not all studies that examined recent homelessness as a potential predictor for clustering have observed a statistically significant association (Moro, 2002; Diel, 2002), none have observed a protective association. In all U.S. genotyped cases from 2005-2009, Moonan and co-authors observed that being homeless within the last 12 months was significantly associated with clustering with an aOR of 1.4 (99% C.I. 1.2-1.6) (2012). In an ethnically diverse region of southern California, Rodwell and colleagues found an aOR of 3.44 (95% CI 1.65-7.18) for homelessness (2012). In Slovenia, Zolnir-Dovc and colleagues also observed that a history of homelessness was significantly associated with clustering with an aOR of 5.73 (95% C.I. 1.21-27.13) (2003). In England, Love and co-authors found an aOR of 5.5 (95% C.I. 1.2-24.1) for homelessness and clustering (2009).

The reasons postulated for homeless persons being at greater risk for clustering vary but one reason may be that homeless people congregate in environments where TB cases and outbreaks often occur such as homeless shelters or homeless 'camps' (McElroy, 2003; CDC, 2003; CDC, 2005). Additionally, this population may also have delay in seeking medical care for diagnosis and treatment resulting in longer periods of infectiousness and transmission (Tan de Bibiana, 2011; McAdam, 2009).

Homeless Shelter (current, or within prior 12 months). A few studies have examined stay in a homeless shelter as a potential predictor, either as a subset of the homelessness covariate or separate from homelessness. Similar to homelessness, persons with a recent history of stay in a homeless shelter may be at higher risk for clustering (Barnes, 1997; Malakmadze, 2005). In Los Angeles, Barnes and co-authors

observed that compared with non-clustered patients, patients from clusters were significantly more likely to have spent time at 3 shelters and other locations when at least 1 patient in the cluster was contagious, and these locations were independent predictors for clustering (1997). Barnes also noted that among non-homeless persons, clustered patients were significantly more likely than non-clustered patients to have used daytime services at any of three shelters (Barnes, 1997). Additionally, in a Wisconsin study investigators observed that clustered cases were associated with homelessness, residence in homeless shelters and single room occupancy hotels. This apparent increased risk for clustering may actually be just better detection. Because homeless shelters have been the site of several TB outbreaks in the U.S. (Barnes, 1997; McElroy, 2003; CDC, 2003; CDC, 2005), many homeless shelters perform routine PPD's on residents. This intervention may result in better case finding, discovering small clusters of two or three persons instead of the singleton cases that might have been discovered outside the shelter environment where their genotype match (es) may go undetected.

Incarceration (current, or within prior 12 months). A couple of studies have found that being incarcerated in prison or jail is an independent predictor for clustering (Moro, 2002, Kempf, 2005). Kempf and co-authors observed an aOR of 2.9 (95% C.I.1.3-6.6) (2005) for the association of clustering and residence in a correctional facility in the year prior to diagnosis. In an Italian study Moro and co-authors observed that aOR of 2.03 (95% C.I. 1.41-2.92) among non-AIDS patients, but did not observe a significant association among AIDS patients (Moro, 2002). Importantly, Moonan's U.S. wide population-based

study (previously cited) did not observe an association with residence in a correctional facility at the time of diagnosis (aOR 0.8; 99% C.I. 0.7-1.0) (2012). However, Moonan did not examine incarceration within the year prior to diagnosis, only current incarceration.

The reasons for this observed association are several. Prisons and jails are locations where TB outbreaks occur often, cases go unrecognized, and medical care is suboptimal. TB transmission is facilitated by crowding, delay in diagnosis, treatment non-compliance, improper isolation; healthy prisoners are mixed with unhealthy prisoners, and questionable nutrition (Moro, 2002; Kempf, 2005). This population, especially in jails is transient. In jails, PPDs are routinely placed on new inmates, but may never be read. Thus, cases may be missed and later diagnosed on the outside.

Alcohol (current, or history of abuse.) Several studies have found that alcohol abuse is one of the strongest risk factors for clustering. In Arkansas, rates of clustering were 19.8% among those with a history of alcohol abuse versus 8.6% among those without (*p-value* < 0.0001) (Talarico, 2011). Also, Ellis and Kempf observed similar results among other U.S. populations (2002 & 2005, respectively). Moonan's comprehensive U.S.-wide study, noted earlier, also found a strong association between clustering and being more likely to abuse substances with an aOR of 1.4 (99% C.I. 1.3-1.7), where substance abuse was any abuse of drugs or alcohol (2012). Non-U.S. studies have observed similar results. In Londoners, Maguire and co-investigators found that alcohol dependence was an independent risk factor for clustering with an aOR of 2.33 (99% C.I 1.46-3.72). In Slovenia, cases with a history of alcohol abuse were at a significantly increased risk for clustering with an aOR of 1.88 (95% C.I. 1.10-3.23). In

Germany, alcohol abuse was also strongly associated with clustering (aOR = 5.11, 95% C.I. 2.77–9.43) (Diel, 2002). In Fok's meta-analysis, alcohol abuse was strongly associated with clustering in both low TB incidence populations with a pooled unadjusted and adjusted OR of 2.2 (95% C. I. 1.6–2.9), and high TB incidence populations with a pooled unadjusted and adjusted OR of 1.4 (95% C. 1.1–1.7) (2008).

Persons that abuse alcohol may be at greater risk for clustering for a number of reasons. They may (1) congregate closely in social settings such as bars (2) have poor recall of social contacts and events, thus are not able to provide a comprehensive contact list to public health investigators, (2) have unknown, thus uncontrolled for co-morbidities that increase their vulnerability for disease as well as decrease the time they progress from infection to disease, (3) experience delays in diagnosis and treatment, or (4) be more prone to treatment non-adherence due to the interaction of alcohol and common TB drugs. These four issues may result in alcoholics becoming infectious earlier and staying infectious longer more so than non-alcoholics, and then transmitting to casual social contacts.

Illicit Drugs (current, or history of abuse). Some studies have found that abusing illegal/illicit drugs is an independent predictor for clustering (Ellis, 2002; Driver, 2006). In the Moonan study mentioned in the alcohol section above, where investigators examined a composite variable that combined injection drug use, non-injection drug use and excessive alcohol abuse into one covariate called substance abuse, they observed an aOR of 1.4 (99% C.I. 1.3-1.7) for all substance abuse and clustering (2012). In an earlier study that used the same composite definition as the

Moonan study and examined all U.S. TB patient genotyped isolates from 2004 to 2005, investigators found that among U.S. born patients the odds of involvement in a county-level genotype cluster were 2.3 (99% C.I. 2.0-2.7) for substance abuser versus a non-abuser. The odds of involvement in a county level cluster for foreign-born patient were less, but still significant at 1.5 (99% C.I. 1.2-2.0) among those who reported substance abuse (Oeltmann, 2009). Further, Oetlmann's study observed that those who abuse substances were more likely to have sputum smear positive disease and to experience treatment failure (2009). The reasons postulated for this association are comparable to those for the association of alcohol abuse and clustering. These may include poor contact and social history recall, poor nutrition and other drug abuse related health issues resulting in quicker progression from infection to disease, and delayed diagnosis. Further, persons who abuse alcohol, crack cocaine, heroin, marijuana, and methamphetamines have all been shown to experience significant weakening to one or more important immunologic mechanisms (Gamble, 2006; Baldwin, 1997; Lysle, 2000; Tashkin, 2002; Friedman, 2003; Mahajan, 2006).

In contrast with alcohol abuse, there is the additional issue. Drug addicts often use secretly in covert locations, and may be unlikely to reveal these locations or drug-use contacts to public health investigators due to the illegal nature of this habit. Furthermore, often substance abuse occurs in enclosed spaces with intentionally limited or poor ventilation and high volumes of human traffic likely increasing the odds of TB transmission (Oeltmann, 2006; Oeltmann, 2009). Additionally it has been noted that

persons who abuse substances are less likely to seek treatment resulting in an extended period of infectiousness and advanced disease at diagnosis (CDC, 2003).

Injection Drugs (subset of Illicit Drugs) (current, of history of abuse). Some studies were able to stipulate this relationship by examining non-injection drug users separately from injection drug users. The results of these studies varied. Kempf and co-investigators found that clustered patients had a crude OR of 1.9 (95% CI, 1.1 to 3.3), among non-injection drug users (2005) but no effect for clustering among infection drug users. While Hernandez and co-authors observed that among patients 60 years and younger that were injection drug users the aOR for clustering was 3.0 (95% C.I. 1.4 to 6.7) (2002). Fok's meta-analysis found that injection drug was associated with clustering in the 10 low incidence population-based studies they examined with a pooled unadjusted and adjusted OR of 2.9 (95% C.I. 2.0–4.2) (2008). The reasons for these inconsistent findings among studies are not entirely clear but may represent a difference in the associated drug culture and the high-risk environments among the study populations. For instance among European populations, 'shooting galleries' may be common, where people might congregate and spread disease. On the other hand, this may not be as prevalent in the U.S. where there might be less congregating among injection drug users. While in the U.S. 'crack houses' are places where persons are smoking drugs, congregating, and sharing TB germs which might be why Kempf's U.S. study observed an association among non-injection drug users only. Further, studies have found that prolonged use of many inhaled or smoked drugs principally crack cocaine, leads to increased coughing and other negative pulmonary effects (Leonhardt,

1994) which may facilitate the spread of TB. It is worth nothing that to date, most studies have not examined the drug abuse and clustering relationship to the granularity of injection drug use versus non-injection drug use, but more research is needed.

Clinical

Human Immunodeficiency Virus (HIV) (seropositivity) (AIDS diagnosis). By definition, confirmed TB cases that are HIV positive have AIDS because TB is one of several conditions that coupled with HIV infection is an AIDS-defining condition. Thus, HIV positive TB cases may or may not be aware of their HIV infection before their TB diagnosis. However, once a patient tests positive for HIV concurrently with their TB diagnosis they actually have three new diagnoses: HIV infection, 'full-blown' AIDS and TB disease. For this reason, all HIV positive persons in my study population were considered AIDS patients and were analyzed as part of an HIV seropositivity (inclusive of all HIV and AIDS cases) covariate analysis. Meaning, this analysis did not attempt to distinguish those HIV positive persons that may or may not have been diagnosed with AIDS prior to their TB diagnosis. Most other studies also assess HIV seropositivity in this way and do not perform a subset analysis of AIDS cases or any analysis by immune-competence among HIV infected persons.

A few U.S. studies have observed HIV seropositivity to be an independent predictor for being part of a TB genotype cluster (Ellis, 1994; Talarico, 2011; Moonan, 2012) while others have found no relationship (Kempf, 2005; Rodwell, 2012). Ellis and co-investigators observed a relative risk of 1.37 (95% C.I. 1.29% to 1.46%), *p-value* = <0.001 for being HIV positive and clustering (2002). Small and colleagues found in

patients under 60 years of age an aOR of 1.8 (*p-value* = 0.04) for being HIV positive and clustering (1994). Talarico and co-investigators in their Arkansas study observed that HIV was significantly overrepresented in medium size TB clusters with 4.6% versus 4.0% (*p-value* < 0.0001) (2011). In Moonan's U.S.-wide study investigators observed a crude OR of 1.7 (99% C.I. 1.5–1.9) and aOR of 1.1 (99% C.I. 1.0–1.3) for being HIV positive and TB clustering (2012).

Results of European studies have varied. Moro's Italian study found that 60.2% of AIDS cases were part of a TB cluster (*p-value* < 0.00001) (Moro, 2002). While Franzetti's study of immigrants residing in Italy found no relationship between HIV status and clustering (2010). Also, Samper and colleagues in their Spanish study did not find a relationship between HIV status and clustering (1993). Outside of the U.S. and Europe, no significant relationship is noticeable between being HIV positive and TB clustering particularly in areas with high TB incidence such as Uganda and South Africa (Asiimwe, 2009 & Verver, 2004, respectively). Some investigators hypothesize that this may be because new TB cases in areas of high TB incidence are overwhelming attributed to ongoing community transmission, where the general population is at such high risk for TB regardless of their HIV status, thus HIV status as well as other clinical factors are less likely to be observed to be associated with clustering.

The reasons why this association is inconsistent among studies are not entirely clear. While HIV seropositivity does not necessarily increase the infectiousness of TB cases, it does substantially increase the risk of progression from infection to disease and this progression will happen much quicker than in immunocompetent persons.

Immunocompetent individuals infected with *M. tuberculosis* have approximately a 10% lifetime risk of developing TB (Hopewell, 2000) with half of the risk occurring in the first 1-2 years after infection. However, HIV-infected individuals with latent TB are about 20-30 times more likely to develop TB disease than those without HIV, or at a rate of 8-10% per year (Daley, 1992). It has also been observed that in some outbreak settings, 35-40% of HIV-infected patients exposed to TB in health care or residential settings developed active TB disease within 60-100 days of exposure (DiPerri, 1989; Daley, 1992). The speed and efficiency by which HIV infected persons develop active disease may be the primary reason why they have been observed to be at higher risk for clustering in some studies. Although why this positive association has been observed in some U.S. studies (Ellis, 1994; Talarico, 2011; Moonan, 2012) and not others (Kempf, 2005; Rodwell, 2012) is unclear, but it may be related to the heterogeneity of the HIV infected population under study. For instance, those immunocompetent HIV infected persons (i.e. those people being effectively treated for their HIV and/or in early HIV infection with high CD4 counts) probably do not experience quicker progression from TB infection to disease while those who are immunocompromised do experience quicker progression to TB disease.

Previous TB Diagnosis (yes). Very few studies on risk of TB clustering have examined previous TB diagnosis as a risk factor. Most investigators choose to examine previous TB treatment instead because previous TB treatment is associated with both clustering and MDR-TB. Unfortunately, in my dataset I did not have complete and verifiable information on previous TB treatment for many of my cases. Thus, this

investigation examined previous TB diagnosis as a risk factor because it had the least missing data. Previous TB diagnosis may be interpreted a few different ways. It could mean (1) previously diagnosed untreated TB, (2) previously diagnosed treated TB, (3) previously diagnosed TB occurring long enough in the past to be considered a separate incidence or activation, and (4) or recent TB diagnosis outside the area of the current public health system. However, according to the RVCT training manual a previous TB diagnosis has a very specific case definition. A patient is considered to have had a previous diagnosis of TB disease if TB disease was verified in the past or the patient completed therapy for TB disease (even if the case-to-case interval is within 12 months); or the patient with TB disease was lost to supervision for more than 12 months and now has verified TB disease again. The RVCT also notes that recurrent cases within 12 months of completion of therapy should be considered previous diagnoses regardless of whether the initial and the subsequent genotypes are the same or are different (CDC, June 2009). The RVCT further emphasizes that written documentation of the previous episode of TB disease is ideal. Nevertheless, states that if the TB disease episode occurred years ago or in another location (e.g., other country); oral report of a previous episode of TB disease is acceptable only when written documentation is not available. Due to the strictness of the previous TB diagnosis case definition and verification, I suspect this covariate to be of high validity in my sample.

In one recent study that examined previous TB, diagnosis the population was Londoners. Hamblion and co-authors found previous TB diagnosis to be significantly associated with clustering with an aOR of 2.1 (95% C. I. 1.5–3.0) (Hamblion, 2016). It

is also worthwhile to discuss studies that have examined prior TB treatment as a risk factor because prior TB treatment is also used to define prior TB diagnosis (on the RVCT). A few studies have observed that having received previous TB treatment is a risk factor for clustering. A South African study compared cases that were retreated for TB after non-compliance from previous treatment compared to the reference group of a newly treated cases and found an aOR of 2.36 (95% C.I. 1.08-5.13) for clustering (Verver, 2004). In England, Love and co-authors observed an aOR of 3.7 (95% C.I. 2.2-6.5) for risk of clustering when comparing retreated cases versus new cases. Likewise, Chan-Yeung and colleagues found an aOR 6.12 (95% C.I. 1.82-20.5) for clustering when comparing retreated cases to new cases (Chan-Yeung, 2006). It is not entirely clear why this may be occurring. Most U.S. studies have either not examined this relationship or have not found an association. However, some studies have examined previous TB diagnosis and found this to be a risk factor for clustering. Previous treatment for TB may be an indicator of patient non-compliance for current treatment as persons that have relapsing TB are likely to have this due to treatment failure or non-compliance. Treatment non-compliance and relapsing TB facilitate the spread even among communities with comprehensive TB control programs.

Rates of Clustering and Recent Transmission

Because the literature suggests that in population-based studies TB isolates sharing identical genotype profiles (also known as clustered isolates) are likely a result of recently acquired infection (Driver, 2006; Ellis, 2002) many of these studies approximate what proportion of TB in their population may be due to recent transmission from their

cluster analysis. Because the proportion of recent TB transmission is a reflection of the success of control measures, correctly accessing this is of public health importance. Research Question 1(b) approximated recent transmission for my study population. Studies have found varying rates of TB clustering and recent transmission among different populations. Some clustering studies do not estimate the rate of recent transmission from their clustering rates due to limitations of their study design. Table 1 summarizes these rates.

Table 1

Rates of Tuberculosis Clustering and Recent Transmission Observed in Previous Population-Based Studies

Study (First Author, Publication Year, Location)	Time Frame	Cases (N)	Clustering (%)	Recent Transmission (%)
Small (1994), San Francisco	1991-92	473	40.0	not estimated
Samper (1997), Zaragoza (Spain)	1993	226	39.0	not estimated
Barnes (1997), Central Los Angeles	1994-96	162	59.0	not estimated
van Soolingen (1999), The Netherlands	1993-97	4,266	46.2	not estimated
Verver (2001), South Africa	1993-98	797	72.0	not estimated
Moro (2002), Milan (Italy)	1995-97	581	41.1	28.1
Kulaga (2002), Montreal	1997-98	243	7.0	4.0
Diel (2002), Hamburg (Germany)	1997-99	423	33.9	20.6
Ellis (2002), U.S. 7 Sentinel States	1996-00	10,752	48.0	not estimated
Hernandez-Garduno (2002), Vancouver	1996-00	793	17.3	not estimated
Weis (2002), Tarrant County (Texas)	1995-96	159	48.0	36.0
Maguire (2003), London	1995-97	2,042	22.7	14.4
Pena (2003), Gran Canaria (Spain)	1993-	145	72.3	58.5

	96			
Zolnir-Dovc (2003), Slovenia	2001	301	37.9	25.0
	1991-03			
Cattamanchi (2005), San Francisco		2,094	18.9	not estimated
	1992-98			
Cacho-Calvo (2005), Madrid (Spain)		448	50.7	not estimated
	1994-00			
Kempf (2005), Alabama		1,834	41.0	35.0
	1999-00			
Chan-Yeung (2006), Hong Kong		702	24.5	15.3
	2001-03			
Driver (2006), New York City		2,408	36.2	27.4
Fok (2007), 17 countries-Meta	1988-	88-		
Analysis	02	10,752	7.0-72.3	not estimated
	2001-04			
Durmaz (2007), Malatya (Turkey)		306	22.0	13.1
Love (2008), England	1998	2,265	16.4	12.2
	1993-00			
Franzetti (2010), Italy		1,999	46.0	not estimated
	1996-03			
Talarico (2011), Arkansas		993	39.5	not estimated
	1990-05			
Kamper-Jorgensen (2012), Denmark		4,601	56.0	not estimated
	2005-08			
Rodwell (2012), Southern California		832	58.0	45.0
	2005-09			
Moonan (2012), United States		36,860	23.1	23.1

Genotyping and Multidrug Resistant TB

One of the important uses of TB genotyping globally is to examine the strains that are most frequently observed as multidrug resistant. The U.S. population has not had extensive spread of multidrug resistant TB thus far. Most cases of MDR-TB in the U.S. have been imported from countries where it is an increasing public health problem. However, when isolated cases are detected in the U.S. it is important to determine the source and the genotype of each case quickly so that local transmission can be ruled out. This section of my literature review will discuss the importance of TB genotyping as a tool to describe and combat MDR-TB.

The European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis conducted a meta-analysis to determine how widespread the Beijing/W genotype of *M. tuberculosis* was, whether it was on the increase, and whether it had a tendency for drug resistance (2006). Individual-level data on >29,000 patients from 49 studies in 35 countries were combined to assess the Beijing genotype's prevalence worldwide and drug resistance. The authors found four patterns for Beijing/W genotype TB, which they described as follows:

1. endemic, not associated with drug resistance at high levels in most of East Asia, lower levels in parts of the US;
2. epidemic, associated with drug resistance at a high levels in Cuba, the former Soviet Union, Vietnam, and South Africa, lower levels in parts of Western Europe;
3. epidemic but drug sensitive in Malawi and Argentina; and
4. very low levels or absent in parts of Europe and Africa.

This essential study established that Beijing/W genotype TB was an emerging in several regions and a predominant endemic strain in others. It was also frequently associated with drug resistance.

A study by Shubladze and colleagues from Georgia (in the former Soviet Union) was conducted on 634 MDR-TB strains were examined for which an MDR phenotype had been previously determined by the proportions method (2013). This study investigated the frequency of major drug resistance mutations across rpoB, katG and

inhA loci of Georgian MDR-TB strains and explored differences between new and previously treated patients. Rifampin resistance was seen in 92.9% of patients and INH resistance was seen in 92.1% of patients; 67.2% and 84.3% of MDR strains harbored respectively rpoB S531L and katG S315T mutations. The inhA C15T mutation was detected in 22.6% of the strains, whereas rpoB H526D, rpoB H526Y, rpoB D516V and inhA T8C were revealed at a markedly lower frequency (≤ 5.2%). The specific mutations responsible for the rifampin resistance of 110 isolates (17.4%) could not be detected as no corresponding mutant probe was indicated in the assay. All types of predominant mutations were observed at higher levels in new cases. Authors concluded that a large portion of Georgian MDR-TB strains have a strong preference for the drug resistance mutations. Further, investigators surmised that MDR TB strains with these mutations might continue to occur in Georgia even in the absence of antibiotic pressure.

A Mexican study by Macías and co-investigators whose goal was to determine the frequency of drug resistance and the clonality of genotype patterns in *M. tuberculosis* clinical isolates from pediatric patients (2011). Resistance to any anti-TB drug was detected in 26.7% of the isolates; 23.3% and 11.1% were resistant to Isoniazid and Rifampicin, respectively, and 11.1% strains were MDR-TB. Spoligotyping produced 55 different patterns; 12/55 corresponded to clustered isolates (n = 47, clustering rate of 52.2%), and 43/55 to unclustered isolates (19 patterns were designated as orphan by the SITVIT2 database). Database comparison led to labeling of 36 shared types (SITs); 32 SITs (n = 65 isolates) matched a previous shared type in SITVIT2, whereas four SITs (n = six isolates) were newly created. Lineage classification based on

principal genetic groups (PGG) showed that 10% of the strains belonged to Bovis and Manu lineages. Among the Bovis and Manu group, the most predominant clade was the Latin-American and Mediterranean (LAM) in 27.8% of isolates, followed by Haarlem and T lineages. The number of single drug-resistant (DR) and MDR-TB isolates in this study was consistent with prior studies in adult populations with risk factors.

Dhatwalia and colleagues conducted a study to determine the prevalence of different genotypes and examine their association with drug resistance among clinical isolates of *M. tuberculosis* from the northern region of India. Investigators analyzed 100 clinical isolates of *M. tuberculosis* using MIRU genotyping and TbD1 analysis. The analysis showed that 34% of strains belonged to the Delhi/CAS lineage, 32% had unknown patterns (27 TbD1-, 5 TbD1+), 18% were of Beijing genotype and 11% were of EAI lineages. Twenty-one strains were MDR-TB, nine of which belonged to the Delhi/CAS lineage, four were of Beijing lineage, six were of unknown pattern and one was of EAI lineage. Their analysis showed the overall proportion of CAS lineage to be 42.96% (95%CI 33-52); the CAS lineage had no association with MDR-TB (OR 0.89, 95%CI 0.66-1.20). This study indicated that the distribution and identification of different genotypes of *M. tuberculosis* could facilitate better understanding of the dynamics that influence disease transmission and drug resistance.

In summary, TB genotyping has been used to successfully describe geographic areas of high TB drug resistance and, in some cases predict continued spread of drug-resistance. These studies underscore the need to continue to develop and expand genotyping capabilities globally as MDR-TB and XDR-TB continue to emerge.

Summary and Conclusions

This literature review provided background and evidence for my conceptual framework, explained the rationale for my choice of statistical analysis, summarized research on the risk factors for clustering, and reviewed clustering and recent transmission rates observed in previous studies. Several population-based molecular epidemiology studies have been conducted to estimate the extent of ongoing TB transmission and to characterize recent transmission dynamics in different geographic regions. Additionally, many of these studies have investigated risk factors for clustering by comparing the characteristics of clustered and nonclustered cases. Some characteristics that studies have consistently found to be associated with being part of a TB genotype cluster were being native-born, being male, younger age, history of alcohol abuse, black race, and history of homelessness. Some characteristics that studies have inconsistently found to be associated with being part of a TB genotype cluster were being HIV positive, having pulmonary disease, and illicit drug use. Better knowledge of transmission risk factors may help to develop more effective prevention strategies to target high-risk populations. The bulk of TB cases in developed countries with comparatively low TB incidence were once thought to be due to the reactivation of infection acquired in the past. Since the advent of molecular epidemiology, TB genotyping studies have strongly suggested that there is greater ongoing transmission and development of active disease than previously appreciated. A better understanding of the factors that influence TB transmission is therefore vital in the global effort to control TB. The purpose of this research was to estimate the proportion of TB in South Carolina that may be due to recent transmission, and determine the risk factors

associated with the genotype clustering of identical *M. tuberculosis* isolates from TB patients in South Carolina from 2005-2011. The social change implication of this study was to further TB elimination by elucidating risk factors for ongoing transmission in South Carolina, thereby informing the South Carolina TB Control Program and other states' control efforts. Chapter 3 will discuss the methodology of the study including the population, sample size, data collection techniques, and data analysis procedures.

Chapter 3: Methodology

In this chapter, I describe the research methodology used in this study, including the rationale for the study design, the setting of the study, the study participants, the instrumentation, the data collection, and the data analysis plan. I also discuss the relevant ethical considerations. The goal of the study was to determine the risk factors associated with genotype clustering of identical *M. tuberculosis* isolates, and to estimate the proportion of South Carolina TB that could be due to recently acquired infection.

Study Design and Rationale

I used a cross-sectional approach to determine what risk factors among incident cases of TB from 2005 to 2011 in South Carolina were significantly associated with being a part of an identical TB genotype cluster of two or more cases. The risk factor assessment for all confirmed cases of TB in South Carolina occurred at the time of TB diagnosis (usually a few weeks prior to the determination of the case's genotype). Because risk factors were queried after TB disease had manifest, the RVCT assessment was considered cross-sectional in nature. However, for the purposes of this study, as has been the case in most studies of this kind, many of the risk factors I examined could be assumed to have preceded TB diagnosis and in some circumstances, TB infection. The risk factors analyzed as covariates were as follows: being U.S. born, being male, being younger, being Black, being Hispanic, history of homelessness, history of incarceration, history of alcohol abuse, history of illicit drug abuse, being HIV positive, and previous TB treatment. Those risk factors for which it cannot be determined whether they

preceded diagnosis are noted (as cross-sectional in nature) and interpreted accordingly in Chapter 5.

Cross-sectional study designs have been used frequently to study risk factors for TB genotype clustering (Oeltmann, 2008; Ricks, 2011; Moonan, 2012). A cross-sectional study design has some advantages, for example, it is a low-cost design and the resulting odds ratios are simple to calculate and easy to interpret (Creswell, 2009). However, cross-sectional designs have limitations. The most important limitation is that they are unable to determine whether the risk factor precedes the outcome of interest. This means that a statistically significant association maybe illuminated between two factors but the temporality of an exposure and outcome cannot be established in a cross-sectional study, and thus causality cannot be established. However, in addition to the low cost and ease of working with the resulting odds ratios, there are two other important points in defense of the cross-sectional design. First, a general point: Even when temporality may be established, such as with the longitudinal cohort design (for instance, one that results in a strongly associated exposure and outcome); this, in and of itself, does not necessarily establish causality. Temporality is but one of several criteria for causality. Bradford Hill's original nine criteria for causality are: (1) strength of association (2) temporality (3) consistency (4) biological plausibility (5) coherence (6) specificity in the causes (7) dose response relationship (8) experimental evidence and (9) analogy. All of these criteria are rarely applied in modern epidemiology, and Rothman and Greenland have been critics of them for reasons they explain in "Causation and Causal Inference in Epidemiology" (Rothman & Greenland, 2005), most schools of Public Health still teach at least five of

them. Those are (1) strength of association (2) temporality (3) consistency (4) biological plausibility and (7) dose response relationship. Thus, even though a longitudinal design is advantageous over a cross-sectional design in that temporality may be established, it may not necessarily establish causality any more than a cross-sectional study. The second point is that many of the risk factors examined in this study may be assumed to have preceded TB diagnosis/disease. Thus, temporality was established due to the nature of the risk factor. Table two lists each covariate examined, whether it could or could not reasonably be assumed to have preceded TB diagnosis, and whether it could or could not reasonably be assumed to have preceded TB disease/infection.

Table 2

Chronology of Potential Risk Factors for Tuberculosis Genotype Clustering

Risk factor	Did this proceed TB diagnosis?	Did this proceed TB disease/ infection?
Age	Yes	Yes
Sex at Birth	Yes	Yes
Race	Yes	Yes
Ethnicity	Yes	Yes
Being U.S. Born	Yes	Yes
HIV Status	Unknown	Unknown
Homeless within the Past Year	Unknown	Unknown
Resident in a Correctional Facility at the Time of Diagnosis	Yes	Unknown
Alcohol Abuse	Yes	Unknown
Drug Abuse	Yes	Unknown
Injection Drug Use	Yes	Unknown
Previous TB Diagnosis	Yes	Yes

Setting of the Study and Population

The setting of this study was the state of South Carolina from 2005 to 2011. The population of South Carolina ranged from 4,012,012 in 2005 to 4,625,401 in 2010. The incidence of TB in South Carolina ranged from 6.1 in 2005 to 3.0 per 100,000 population in 2011 (1,346 TB confirmed cases in those 6 years. All South Carolina residents were considered at risk for TB in the years preceding, during, and after this study. The sampling frame for this study was confirmed cases of TB disease (n = 1,346) excluding all exclusively extrapulmonary TB (n = 266) in South Carolina from 2005 to 2011, for a total sampling frame of 1,080 pulmonary (including pleural) or both pulmonary (including pleural) and extrapulmonary TB disease. This study did not use any sampling procedures and all TB cases for which my dataset contained a valid genotype were considered in the study analysis. The analysis dataset contained a valid genotype for 63% of eligible cases in the sampling frame (n = 685) (see Table 3). This study did not use any sampling procedures for two reasons. First, I wanted to use all genotype data available to maximize statistical power. Second, as mentioned in Chapter 2 Limitations, I did not have complete demographic, clinical, or social information on TB cases for which I did not have a valid genotype. Thus, I did not have an entirely accurate way to verify the randomness or accuracy of my sampling methodology. Table 3 displays South Carolina TB statistics for 2005 through 2011, the associated sampling frame, and sample size per year.

Table 3

The Percentage of Confirmed Tuberculosis Cases in South Carolina with a Genotype Determined, 2005-2011

Year	Confirmed Cases	Case Rate[a]	Eligible for Genotype[b]	Successfully Genotyped	Genotyping Rate[c]
2005	261	6.1	214	126	59%
2006	222	5.1	178	105	59%
2007	218	4.9	175	116	66%
2008	188	4.2	146	95	65%
2009	164	3.6	138	85	62%
2010	153	3.3	121	84	69%
2011	140	3.0	108	74	69%
Total	1,346	3.0-6.1	1,080	685	63%

[a] *Rate per 100,000 population.*
[b] *Cases with pulmonary (including pleural) TB (may have both pulmonary and extrapulmonary).*
[c] *Rate calculated as cases genotyped per year/eligible for genotype.*

Instrumentation

All information on covariates of interest for this study was obtained from the RVCT. There is a copy of the RVCT displayed in Appendix A. The Centers for Disease Control and Prevention National Center for HIV/AIDS, Viral Hepatitis, STD, and TB prevention developed and owns all versions of this form. This form is provided to all states for the purposes of collecting and reporting information to the CDC TB division on all suspect and confirmed TB cases identified in the states. A training manual is provided to state and local health departments on how to complete this form. All South Carolina health department TB nurses are required to attend an in-person training on completing this form. The CDC website states that the form may be used for relevant research purposes, but should be properly bibliographed.

Below, the specific questions on this form that were used for this study are listed in the order in which they appear on the form.

1. Date of Birth (MM/DD/YYYY),
2. Sex at Birth (Male, Female),
3. Race (American Indian or Alaskan Native, Asian: Specify, Black or African American, Native Hawaiian or Other Pacific Islander: Specify, White),
4. Ethnicity (Hispanic or Latino, Not Hispanic or Latino),
5. Country of Birth (U.S. born or born to a parent who is a U.S. citizen, Country of Birth: Specify),
6. Month-Year arrived in the U.S. (MM/YYYY),
7. Site of TB Disease (Pulmonary, Pleural, Lymphatic: Cervical, Lymphatic: Intrathoracic, Lymphatic: Axillary, Lymphatic: Other, Lymphatic: Unknown, Laryngeal, Bone and/or joint, Genitourinary, Meningeal, Peritoneal, Other: Enter Anatomic Code, Site not Stated),
8. HIV Status (Negative, Positive, Indeterminate, Refused, Not offered, Test Done, Results unknown, Unknown),
9. Homeless within the Past Year (No, Yes, Unknown),
10. Resident in a Correctional Facility at the Time of Diagnosis (No, Yes, Unknown, if Yes select one: Federal Prison, State Prison, Local Jail, Juvenile Correction Facility, Other Correctional Facility, Unknown),
11. Alcohol abuse use in the Past Year (No, Yes, Unknown),
12. Injection Drug use in the Past Year (No, Yes, Unknown),

13. Non-injection Drug use in the Past Year (No, Yes, Unknown), and

14. Prior TB disease diagnosis (No, Yes, Unknown).

The literature review found no reliability or validation studies for the RVCT instrument used as part of the original data collection. However, in Table 4, I have listed all the studies on risk factors of TB genotype clustering that have used all or part of the RVCT as their primary data collection/abstraction source. All of these studies were published in reputable peer-reviewed journals. The limitations of the RVCT form were described Chapter 1: Introduction of Study—Limitations and are described further in Chapter 5: Conclusions and Recommendations.

Table 4

Peer-Reviewed Studies Utilizing the Report of a Verified Case of Tuberculosis (RVCT)

First Author (Publication Year)	Study Title	Publication
Driver (2006).	Molecular epidemiology of tuberculosis after declining incidence, New York City, 2001-2003.	Epidemiology & Infection
Driver (2006).	Which patients' factors predict the rate of growth of Mycobacterium tuberculosis cluster in an urban community?	American J of Epi
Kempf (2005).	Long-term molecular strains in Alabama, a state characterized by a largely indigenous, low-risk population.	J of Clinical Micro.
Moonan (2012).	Using genotyping and geospatial scanning to estimate recent Mycobacterium tuberculosis transmission, United States.	Emer Infectious Dis
Oeltmann (2009).	Tuberculosis and substance abuse in the United States, persons acquired prior to entering the U.S., 2005-2009.	

(table continues)

Ricks (2011).	Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the U.S., 2005-2009.	Public Library of Science
Rodwell (2012).	Factors associated with genotype clustering of Mycobacterium tuberculosis isolates in an ethnically diverse region of southern California, United States.	Infect. Genetic Evol
Talarico (2011).	Identification of factors for tuberculosis transmission via an integrated multidisciplinary approach	Tuberculosis

Data Collection

There were no recruitment procedures used in this study because the study is based entirely on secondary data collected through normal TB case investigation and confirmation procedures. The original data was collected as part of a South Carolina TB case investigation. These investigations occurred for all suspected and confirmed cases of TB in South Carolina, which is a requirement of all states receiving federal TB control and prevention funding.

Data Analysis Plan

I constructed the methodology of my study to answer the following research questions.

Research Question 1:

a) Using the mycobacterial interspersed repetitive unit (MIRU) genotyping method, and spoligotyping, for cluster classification of tuberculosis cases in South Carolina, I estimated the proportion of TB cases that were genotyped clustered versus singleton?

b) Estimated the proportion of South Carolina TB cases that may be due to

recently acquired infection the following logic was applied:

I. For genotype clusters of only 2 cases: One case of the cluster was assumed to be a souce case. One case of the cluster was assumed to be the recently infected case.

II. For genotype clusters of 2 or more cases: One case of the cluster was assumed to be the source case. All other matches in the cluster were assumed to be due to recent transmission.

Thus, *C - 1*, were counted as recently transmitted cases, where *C* was the number of identical isolates in the cluster (based on the "*n - 1* method" index described previously).

Research Question 2: Determined the risk factors of genotype clustering among incident South Carolina TB cases from 2005 to 2011 considering the following hypotheses:

H_0 There is no relationship between being foreign-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being foreign-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being foreign-born and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between age and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being male and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between black race and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between hispanic ethnicity and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between residence in a correctional facility at the time of diagnosis and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between homelessness within the past year and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between being HIV positive and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between alcohol abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between illicit drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between injection drug use and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between substance abuse and being part of a TB genotype cluster when controlling for other significant covariates.

H_0 There is no relationship between having prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a1} There is a positive relationship between having prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

H_{a2} There is a negative relationship between having prior TB disease and being part of a TB genotype cluster when controlling for other significant covariates.

My secondary dataset was provided in a Microsoft excel passphrase-encrypted spreadsheet. For security reasons, the passphrase was provided verbally and was not transcribed. For the analysis phase of this study, I imported the excel data into SAS 9.3. SAS 9.3 was used to conduct all analysis. I cleaned the data as necessary by labeling, formatting, and making sure all values were reasonable and valid. As mentioned in Chapter 2, logistic regression is commonly used in cross-sectional studies to model the relationship between a binary dependent variable and one or more explanatory or independent variables (Hosmer & Lemeshow, 2000) and I used it to model my data. I used SAS code to format the genotype variable, which was my dependent variable. I examined how many different clusters were present in the data, the range of cluster size, and what proportion of cases were clustered (had one or more matches) versus singleton. I coded the dependent variable of interest called "clustered status" as a character variable of "0" for singleton cases and "1"for clustered cases. The independent variables were obtained from the questions on the RVCT**.**

Independent variables. The rationale for inclusion of the covariates chosen was that these factors have been found to be statistically significant risk factors of TB genotype clustering in multiple previous studies (Love, 2009; Barnes, 1997; Cacho-Calvo, 2005; Cattamanche, 2006; Fok, 2008; Kempf, 2005; Chan-Yeung, 2006; Driver, 2006, Ellis, 2002). Categorical variables with two levels are coded as "1" for the level that has been observed in the literature to be positively associated with being part of a TB genotype cluster, and "0" for the other level. Country of birth was coded with "0" assigned to foreign-born and "1" assigned to U.S. born. Date of birth was not provided in the dataset, only age categories (so age was modeled as categorical variable rotating both younger and older age groups as the risk factor see Chapter 5 Results for more information). Sex was coded with "0" assigned to females and "1" assigned to males. Race was a categorical variable divided as (1) White (2) Black (3) Hispanic, (4) Asian, and (5) Other (the "Other" category was the collapsed category of American Indian or Alaskan Native and Native Hawaiian or Other Pacific Islander). Homeless within the past year was "0" as no and "1" as yes. Resident in a correctional facility at the time of diagnosis was "0" for no and "1" for yes. HIV status was "0" for negative and "1" for positive. Prior or current alcohol abuse was "0" for no and "1" for yes. Prior or current drug abuse was "0" for no and "1" for yes. Prior TB diagnosis was coded as "0" for no and "1" for yes.

By definition, confirmed TB cases that are HIV positive have AIDS because TB is one of several conditions that coupled with HIV infection is an AIDS-defining condition. Thus, TB cases may or may not be aware of their HIV infection before their

TB diagnosis. However, once a patient tests positive for HIV concurrently with their TB diagnosis they actually have three new diagnoses: HIV infection, 'full-blown' AIDS, and TB disease. For this reason, all HIV positive people in my study population were considered AIDS patients and analyzed as part of an HIV seropositivity, inclusive of all HIV and AIDS cases, covariate analysis. This analysis did not attempt to distinguish those HIV positive persons that may or may not have been diagnosed with AIDS prior to their TB diagnosis. Most other studies also assess HIV seropositivity in this way and do not perform a subset analysis of AIDS cases alone or any analysis by immune-competence among HIV infected persons.

This study used a bivariate and multiple regression statistical analysis. This approach made it possible both clarification of bivariate relationships, as well as a determination of the degree of relationship between and among variables (Kleinbuam, 1988). A multiple logistic regression approach was appropriate for this study because it allowed simultaneous consideration of multiple predictor variables for a binary outcome variable, and subsequent stepwise model building (Hosmer, 1989). First, the bivariate results of all risk factors were examined. Then multiple logistic regression analysis and a stepwise modeling building process included variables with an alpha level of 0.10. For a covariate to be included in the final model, it was statistically significant with a *p-value* < 0.05. The odds ratio results from this analysis were examined and interpreted in the context of the bivariate and multivariable analysis. Model fit statistics were examined and interpreted. As is customary, odds ratios >1.0 with an associated *p-value*

< 0.05 were interpreted as a risk factor for clustering. Likewise, odds ratios < 1.0 with an associated *p-value* < 0.05 were interpreted as protective for clustering.

Ethical Considerations

For this study, I obtained Institutional Review Board (IRB) approval through Walden University on March 17, 2017 (3-17-17-0080120). As a full-time employee of the South Carolina DHEC Division of Infectious Disease Control, they provided me patient-level identifiable data in the summer of 2012 by South Carolina DHEC. This data was not available to the public. I was allowed to view this data as part of my job and it was covered under the confidentiality agreement I had to sign and adhere to as an employee. Since my departure from South Carolina DHEC, I have been covered under the volunteer use agreement as a non-employee.

At the time of my employment, one identifiable passphrase-encrypted dataset was kept on the South Carolina DHEC desktop computer for which I alone had passphrase-protected access. During the time of my employment, South Carolina DHEC procedures for patient data protection were strictly followed including that identifiable data could never be emailed, and it could only be shared with disease investigation staff who had appropriate access rights. At the time just prior to my separation from South Carolina DHEC, I de-identified all data that was to be used in this study. I de-identified the data using the following procedures:

1. I created a randomly generated unique identifier for all subjects.

2. I deleted all personal identifiable variables for each subject including first name, last name, South Carolina TB case number, CDC TB case number, and patient address including city, street, and zip.
3. I created an alias variable for HIV status. The term "HIV" or "AIDS" was not included in any of the data I am using.
4. I deleted all variables except those pertinent to my analysis including any variables or opened-ended comments that included the names of HIV related medications.

The one original copy of the data I was provided was completely deleted and I have had no access to identifiable data since my departure from South Carolina DHEC in October of 2012. Per my volunteer data-use agreement (DUA) with South Carolina DHEC, a copy of the de-identified data (described above) was placed on an encrypted-password-protected jump drive. That data was then placed on my personal laptop in an encrypted password protected file and the jump drive file was deleted. Entry to my personal laptop was password protected and I was the only person that had access to this data.

Informed consent was not obtained from any individual for this study for the following reasons. (1) Inclusion of a subjects de-identified data did not pose any physical or mental risk to the subjects. Any risks associated with primary data collection (such as obtaining a TB specimen for diagnosis and genotyping, or TB treatment) would have occurred as a part of the normal TB diagnosis and treatment procedures. The low to moderate risks associated with TB diagnosis and treatment far outweigh the risks of

having untreated and undiagnosed TB disease for the individual. Further, TB investigation, diagnosis, treatment, and containment are required by *South Carolina State Law Title 44 – Health, Chapter 31 Tuberculosis, and Article 1*. Any use of a patient's secondary data in no way contributed to those risks. (2) Data was provided to only myself for the purposes of this study 2 to 8 years after it was originally collected, and all identifiers were dropped from the data, thus subjects were not identifiable for which to obtain informed consent. All descriptive data and results are displayed at an aggregate summary level in this report and there was no risk of individual identification.

Summary

In this chapter, I discussed the rational for the study design, setting of the study, study participants, instrumentation, data collection, data analysis plan, and any ethical considerations. Cross-sectional study designs have been used in previous retrospective population-based TB genotyping studies, and this was a reasonable and appropriate approach for my study. Each research question was described. Bivariate and multiple logistic regression analysis were used to calculate odds ratios for each covariate examined in this study.

Chapter 4: Data Analysis

The purpose of this research was to estimate the proportion of South Carolina TB that may be due to recently acquired infection, and determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates from TB patients in South Carolina from 2005-2011 by using a multivariable logistic regression technique to model risk factors for the binary outcome of being part of a TB genotype cluster, yes or no. In this study, two research questions were tested by using appropriate statistical methods. This chapter presents an overview of the results and findings of these analyses.

Data Collection and Study Participants

I used a cross-sectional approach to determine the risk factors among incident cases of TB from 2005 to 2011 in South Carolina that were significantly associated with being a part of an identical TB genotype cluster of two or more cases. This study was based entirely on secondary data collected through normal TB case investigation and confirmation procedures. The incidence of TB in South Carolina ranged from 6.1 per 100,000 population in 2005 to 3.0 per 100,000 population in 2011, or 1,346 confirmed TB cases over those 6 years. The sampling frame for this study was confirmed cases of TB disease (n = 1,346), excluding all exclusively extrapulmonary TB (n = 266) in South Carolina from 2005 to 2011, for a total sampling frame of 1,080 pulmonary (including pleural) or both pulmonary (including pleural) and extrapulmonary TB disease. The analysis dataset contained a valid genotype for 63% (685/1,080) of cases in the sampling frame (see Table 3), or 50.9% (685/1,346) of all confirmed South Carolina TB cases from

2005 to 2011. The study population was 685 confirmed cases of TB for which a valid genotype was available. Table 5 displays South Carolina TB statistics from 2005 through 2011, including the associated sampling frame and sample size per year.1).

Table 5

The Percentage of Confirmed Tuberculosis Cases in South Carolina with a Genotype Determined, 2005-2011

Year	Confirmed Cases	Case Rate[a]	Eligible for Genotype[b]	Successfully Genotyped	Genotyping Rate[c]
2005	261	6.1	214	126	59%
2006	222	5.1	178	105	59%
2007	218	4.9	175	116	66%
2008	188	4.2	146	95	65%
2009	164	3.6	138	85	62%
2010	153	3.3	121	84	69%
2011	140	3.0	108	74	69%
Total	1,346	3.0-6.1	1,080	685	63%

[a] *Rate per 100,000 population.*
[b] *Cases with pulmonary (including pleural) TB (may have both pulmonary and extrapulmonary).*
[c] *Rate calculated as cases genotyped per year/eligible for genotype.*

In order reveal any bias that may have been introduced by not using a sampling methodology, I compared the distribution of covariates in the cases that were genotyped (the 685 cases of the study population) to all confirmed TB cases in South Carolina (the 1,346 confirmed TB cases) from 2005-2011. This was done to determine how closely the genotyped sample resembled the entire case population. This is discussed in Results, under Estimate of Recent Transmission.

Descriptive Statistics

Table 6 below describes the characteristics of the study population. Most cases in the study sample were adults with 31% being between 25-44 years, 38% 45-64 years, and

20% 65+ years old. Children under 15 years made up less than 1% of the sample. The majority of the population was male at 67.6 %. Sixty-three percent of the sample was Black, 16% White, 13% Hispanic, and 8% other races. Close to 80% of the cases were US-born. Less than 9% were HIV positive. A little under 7% reported having been homeless in the last year, and a little over 2% were incarcerated at the time of diagnosis. Almost 29% of the study population were heavy drinkers, while over 15% reported illicit drug use. Most of the drug users reported non-injection drug use, with only 1.3% affirming injection drug use. Eleven percent of the sample were both alcohol and drug abusers which is termed "substance abuse" for the purposes of this analysis. Lastly, for 96% of the study population this represented their first TB diagnosis, while 4% noted a previous TB diagnosis.

Table 6

Summary of Descriptive Statistics for South Carolina Genotyped Tuberculosis Cases, 2005-2011

	Genotyped Cases *(N=685)*	
Variables	*N*	%
Age (Years)		
00-04	3	0.44
05-14	1	0.15
15-24	75	10.95
25-44	215	31.39
45-64	259	37.81
65+	132	19.27
Age-Collapsed		
Child (0 - 14 years)	4	0.58
Adult (15 - 64 years)	549	80.15
Elderly (65+)	132	19.27
Sex at Birth		
Male	463	67.59
Female	222	32.41

(table continues)

Race/Ethnicity		
Black	432	63.07
Hispanic/Latino	88	12.85
Asian American	50	7.3
White	109	15.91
Other (American Indian, Alaskan Native, Native Hawaiian or other Pacific Islander)	6	0.87
Origin		
US-born	545	79.56
Foreign-born	140	20.44
HIV Status		
Negative	571	83.36
Positive	60	8.76
Unknown (Refused, Not offered, or Missing)	54	7.88
Homelessness		
Yes	45	6.57
No	638	93.14
Unknown	2	0.029
Correctional Institution		
Yes	15	2.19
No	670	97.81
Alcohol		
Yes	195	28.47
No	485	70.08
Unknown	5	0.73
Illicit Drug Use		
Yes	104	15.18
No	574	83.8
Unknown	7	1.02
Non-Injection Drug Use		
Yes	103	15.04
No	575	83.94
Unknown	7	1.02
Injection Drug Use		
Yes	9	1.31
No	669	97.66
Unknown	7	1.02
Substance Abuse		
Yes (yes to alcohol and drugs)	79	11.5
No (no to either alcohol, drugs, or both)	601	87.7
Unknown	5	0.73
Previous TB Diagnosis		
Yes	29	4.23
No	656	95.8

(table continues)

As shown in Table 7, the genotype lineages that occurred in South Carolina TB patients between 2005 and 2011 were mostly EuroAmerican and East Asian with a few IndoOceanic, East African Indian, Bovis type, and Africanum type. This study revealed a modest diversity of circulating genotypes as would be expected in a low incidence region. All of the genotypes are fairly common in the US.

Table 7

Distribution of Genotype Lineages among South Carolina Tuberculosis Cases, 2005-2011

	Genotyped Cases (N=685)	
Genotype Lineage	*N*	%
Africanum type	2	0.29
Bovis type	3	0.44
East African Indian	4	0.58
East Asian	74	10.8
EuroAmerican	528	77.1
IndoOceanic	31	4.53
Unknown	43	6.3

Results

Estimate of Recent Transmission

There were 1,346 confirmed cases of TB in South Carolina from 2005 to 2011. Of those, a genotype was successfully obtained on 685. For this study, research question 1(a) asked *what proportion of TB cases in South Carolina are clustered versus singleton?* Of these 685 cases that were successfully genotyped, 419 were clustered, for a clustering rate of 61.2% (419/685). Research question 1(b) asked *what proportion of South Carolina TB cases might be due to recently acquired infection?* There were 76 different clusters represented by the 419 clustered cases. Cluster size ranged from two2 to 39

cases. As described in Chapter 3 Methodology, for genotype clusters of only two cases, one case of the cluster was assumed a source case while one case of the cluster was assumed the recently infected case. For genotype clusters of two or more cases, one case of the cluster was assumed the source case and all other matches in the cluster were assumed to be due to recent transmission. This method is often called the "*n – 1 method*" and has been used in previous research to estimate recent transmission (Reza Allahyar Torkaman, 2014; Ricks, 2009). This logic provided the calculation:

419 (clustered cases) – 76 (source cases) = 343 recently transmitted cases

This gives a recent transmission estimate of 343/685 or 50.1% for this study population.

Table 8 below displays the descriptive statistics of all TB cases compared to only the genotyped cases. Chi-square analysis was performed to compare all South Carolina TB from 2005 to 2011 to those cases that were selectively in 10 characteristics of interest. For most characteristics, the genotyped cases were not significantly different from the entire South Carolina case population. They were different in three characteristics: age, non-injection drug use, and alcohol abuse. Meaning it appears from this analysis that middle age and older cases were more likely to be selected for genotyping, abusers of non-injections drugs were more likely to be selected for genotyping, and alcoholics were more likely to be selected for genotyping. However, other than these three characteristics, given that the study population was not randomly sampled (as discussed previously this was a convenience sample that included all cases with a genotype for maximum statistical power), this appears to be a reasonable representation of the South Carolina's TB case population from 2005 to 2011, with the exception of age. It is not surprising that

our genotyped sample would not be representative in age. As we discussed previously pediatric cases are not routinely genotyped because positive cultures on children are very difficult to obtain. However, the reasonable similarity between our study sample and our entire case population gives us better confidence that our estimate of recent transmission of 50.1% is generalizable to 2005 to 2011 South Carolina TB cases. The limitations of this estimate will be discussed more in the *Limitations of the Study* section below.

Table 8

Summary of Descriptive Statistics for all South Carolina Confirmed Tuberculosis Cases Compared to all TB Genotyped Cases, 2005-2011

	All Cases *(N=1,346)*		Genotyped Cases *(N=685)*		Pearson's R (*p-value*) All Cases vs. Genotyped Cases	Clustered Cases *(N=419)*	
Variables	*N*	%	*N*	%		*N*	%
Age							
00-04	69	5.13	3	0.44		3	0.72
05-14	33	2.45	1	0.15		1	0.24
15-24	137	10.20	75	10.95		36	8.6
25-44	385	28.60	215	31.39		135	32.2
45-64	453	33.70	259	37.81		172	41.1
65+	269	20.00	132	19.27	0.985 (< 0.00001)	72	17.2
Sex at Birth							
Male	863	64.10	463	67.59		286	68.3
Female	483	35.90	222	32.41	0.992 (0.018)	133	31.7
Race/Ethnicity							
Black	799	59.36	432	63.07			
Hispanic/Latino	196	14.56	88	12.85		25	6
Asian American	104	7.73	50	7.3		18	4.3
White	230	17.09	109	15.91		61	14.6
Other (American Indian, Alaskan Native, Native Hawaiian or other Pacific Islander)	17	1.26	6	0.87	0.999 (< 0.00001)	4	0.24
Origin							
US-born	1,061	78.80	545	79.56		379	90.5
Foreign-born	285	21.20	140	20.44	0.997 (0.003)	40	9.6
HIV Status							
Negative	1,066	79.20	571	83.36		344	82.1

(table continues)

Positive	103	7.65	60	8.76		41	9.8
Unknown (Refused, Not offered, or Missing)	177	13.15	54	7.88	0.993 (< 0.00001)	34	8.1
Homelessness							
Yes	69	5.04	45	6.57		33	7.9
No	1,275	93.13	638	93.14		386	92.1
Unknown	25	1.83	2	0.029	0.997 (< 0.00001)	0	0
Correctional Institution							
Yes	27	2.01	15	2.19		11	2.6
No	1,319	97.99	670	97.81	0.992 (0.008)	408	97.4
Alcohol							
Yes	296	21.99	195	28.47		140	33.4
No	1,038	77.12	485	70.08		275	65.6
Unknown	12	0.89	5	0.73	0.997 (< 0.00001)	4	0.95
Illicit Drug Use							
Yes	172	12.80	104	15.18		77	18.6
No	1,160	86.20	574	83.8		342	81.4
Unknown	14	1.04	7	1.02	0.997 (< 0.00001)	0	0
Non-Injection Drug Use							
Yes	157	11.66	103	15.04		77	18.4
No	1,175	87.30	575	83.94		337	80.4
Unknown	14	1.04	7	1.02	0.997 (< 0.00001)	5	1.2
Injection Drug Use							
Yes	15	1.11	9	1.31		5	1.2
No	1,317	97.85	669	97.66		409	97.6
Unknown	14	1.04	7	1.02	0.997 (< 0.00001)	5	1.2
Previous TB Diagnosis							
Yes	59	4.38	29	4.23		*(table continues)*	
No	1,287	95.62	656	95.77	0.997 (0.003)	[illegible]	[illegible]
					0.995 (< 0.00001)		

Characteristics of Clustered Cases

Table 9, below, displays the characteristics of non-clustered versus clustered cases. For clustered cases, 68% were male, 74% were Black, and over 90% were US-born. Over 90% were over 25 years old. Just under 6% had a previous TB diagnosis. Over 33% were alcohol abusers and about 18% were drug users. Less than 3% were

incarcerated at the time of TB diagnosis. Less than 10% were recently homelessness, and less than 10% were HIV positive. Based on the chi-square test results reported in Table 9 below, clustered cases were statistically more likely to be U.S. born, black, alcohol abusers, abusers of non-injection drugs, and of older age.

Table 9

Characteristics of South Carolina TB Genotype Clustered and Non-Clustered Cases, 2005-2011

N=627, Clustered (n=382), Not Clustered (n=245)

Variables	Point Estimate	95% Wald C.I.	Wald X^2	*p-value*
Age				
Child (0 - 14 years) vs. Elderly				
Adult (15 - 64 years) vs. Elderly				
Old (25-65+years) vs. Young (0-24 years)	1.770	1.09-2.86	5.3454	0.0208
Sex at Birth				
Male vs. Female	1.060	0.75-1.49	0.106	0.7447
Race/Ethnicity (White = reference group)				
Black vs. White	2.060	1.29-3.28	51.58	<0.0001
Hispanic/Latino vs. White	0.329	0.18-0.61	18.67	<0.0001
Other vs. White	0.550	0.28-1.09	2.41	0.121
Origin				
US-born vs. Foreign-born	5.570	3.66-8.47	64.47	<0.0001
HIV Status				
Positive vs. Negative	1.390	0.78-2.50	1.2322	0.267
Homelessness				
Yes vs. No	2.010	0.99-4.06	3.8052	0.0511
Correctional Institution				
Yes vs. No	1.790	0.56-5.67	0.9666	0.3255

(table continues)

Alcohol				
Yes vs. No	2.030	1.40-2.94	13.93	0.0002
Illicit Drug Use				
Yes vs. No	2.060	1.28-3.32	8.7344	0.0031
Non-Injection Drug Use				
Yes vs. No	2.150	1.32-3.50	9.5823	0.002
Injection Drug Use				
Yes vs. No	0.799	0.21-3.01	0.1101	0.74
Substance Abuse				
Yes vs. No	2.260	1.30-3.93	8.26	0.004
Previous TB Diagnosis				
Yes vs. No	2.320	0.92-5.83	3.192	0.074

(table continues)

Bivariate Analysis

The second research question asks what are the risk factors for genotype clustering among incident South Carolina TB cases from 2005 to 2011? This question was first examined with a bivariate unadjusted analysis. As stated previously, the initial sample size was 685 genotyped cases. For the purposes of all the logistic analyses, I dropped cases for which I did not have complete information on all covariates. There were 54 cases with unknown HIV status, 1 of which also had unknown drug and alcohol history. There were 2 cases with unknown homeless status, 1 of which also had unknown drug history and HIV status, and the other had unknown drug and alcohol history. When substracting the cases dropped for unknown homeless history, HIV status, and alcohol status there were 2 additional cases with unknown drug status. This resulted in 58 cases with missing information that were dropped from our sample giving an analysis sample of 627 cases for all logistic models; 382 clustered cases and 245 singleton cases. A logistic analysis of the 58 cases dropped for missing data was performed to determine if they were more or less likely to be clustered than the cases not dropped. The results of this analysis are below in Table 10. No significant association between being dropped and being clustered was observed. Thus, we can conclude there was no indication that dropped cases were more or less likely to be clustered (or that clustered cases were more or less likely to be dropped).

Table 10

Dropped Data Analysis

		Not Clustered	Clustered	
Not Dropped	Frequency	245	382	627
	Percent	35.8	55.8	91.5
Dropped	Frequency	21	37	58
	Percent	3.1	5.4	8.5
Total	Frequency	266	419	685
	Percent	38.8	61.2	100
Dropped vs. Not Dropped	*Point Estimate*	*95% Wald C.I.*	*Wald* X^2	*p-value*
	1.130	0.65-2.0	0.184	0.668

As noted earlier, the age variable was only available in the data as a categorical variable (I was not provided date of births in my dataset). Its original categories (as used by the RVCT) were 0-4, 5-14, 15-24, 25-44, 45-64, and 65+ years. When the crude model of age (in these original categories) and clustering was examined, no associations were observed. I decided to collapse the two youngest age groups into 0-15 years as "child" because there were so few children in the sample. This was used as the reference group. The 3 middle age groups were collapsed into 24-64 years and called "adult", and 65+ years was called "elderly". With these new classifications, there was still no association noted for clustering for either child versus elderly or adult versus elderly. Recall from Chapter 2 Literature Review, there was a lot of variation with regard to classification of age groups in prior studies, as well as variation in what ages were found to be at risk for clustering. Additionally, because my sample had very few children ($n = 4$) it seemed prudent to further collapse these categories to see if any association could be detected with larger groups of wider age ranges. Thus, I further collapsed my age categories and tested

the bivariate association between risk of clustering and age where “young” was the reference group (0-44 years) and old was the risk group (45-65+ years). Table 7 below shows that this association was statistically significant with an unadjusted OR of 1.77 (95% C.I. 1.09-2.86) or people over 44 years old were about 1.8 times more likely to be part of a TB genotype cluster than people 44 and younger.

The next crude model examined was sex and TB clustering. Although much of the literature indicates men may be at higher risk for clustering, no association was observed in my population between male sex and risk of clustering, or female sex and risk of clustering. Further, neither of these were significantly protective for clustering.

The next demographic variable examined was the bridged race/ethnicity variable. Although the RVCT does actually inquire about both race and ethnicity as two separate questions, the dataset I was provided only contained this information as a bridged variable (Meaning the data did not allow me to discern non-Hispanic White or Black persons or Hispanic White or Black persons. The only categories available to me were Black, Hispanic, Other, and White). When this variable was examined using White as the reference group (as is done most often in the literature), people of Black race were observed to be at higher risk for clustering with a crude OR of 2.06 (95% C.I. 1.29-3.28). Interestingly, Hispanics when compared to Whites displayed a statistically significant protective effect for clustering, with a crude OR of 0.33 (95% C.I. 0.18-0.61). Based on the finding that Blacks and Whites were at significantly higher risk for clustering when compared to Hispanics, I became suspicious there may be an interaction occurring

between origin (US-born versus Foreign-born) and race in this population. I explored this possibility in the multivariable stage of this analysis discussed in the next section.

The next potential predictor variable that was examined was origin classified as US-born versus Foreign-born. In the crude analysis, being US-born was highly predictive of genotype clustering. Specifically, U.S. born TB cases had over 5.5 times the odds of being part of a genotype cluster than did Foreign-born cases in the unadjusted analysis (OR = 5.57, 95% C. I. 3.66-8.57). In fact, in the unadjusted analysis US-birth was the greatest risk factor for clustering.

Next I examined the crude model of HIV status and clustering. Despite what previous U.S. studies have observed, the bivariate analysis did not show being HIV positive as a risk factor for clustering in this study population. Further, for the crude model of living in a correctional facility at the time of diagnosis and clustering, again despite the positive association some previous studies have observed, no association was observed in my study population. However, being homeless or having been homeless in the year prior to diagnosis was significant enough (at an alpha of 0.10) to be included in the multivariate stage of the analysis with a crude OR = 2.01, 95% C.I. 0.99-4.06. Additionally, alcohol abuse in the year prior to diagnosis was significantly associated with clustering with a crude OR of 2.03 (95% C.I. 1.40-2.94). In addition, drug abuse was significantly associated with clustering with a crude OR of 2.06 (95% 1.28-3.32). When drug abuse was divided into injection drug-users and non-injection drug users, the association remained significant only among non-injections drug users, and in fact, it slightly increased. I found there was a lot overlap between drug and alcohol users so I decided to

combine this into a single covariate called "substance abuse" as has been done in other U.S. studies (Moonan, 2012). Substance abuse (alcohol and drug abusers inclusive) became an even more significant predictor in the unadjusted model with an OR of 2.26 (95% C.I. 1.30-3.93) than either alcohol or drug, abuse was alone. Lastly, having had a prior diagnosis of TB was significant enough to be included in the multivariate stage of the modeling with a crude OR of 2.32 (*p-value* = 0.074). All bivariate results are summarized in Table 11.

Table 11

Bivariate Analysis of the Association between Tuberculosis Genotype Clustering and Covariates of Interest

N = 627, Clustered (*n* = 382), Not Clustered (*n* = 245)

Variables	Point Estimate	95% Wald C.I.	Wald X^2	*p-value*
Age				
Child (0 - 14 years) vs. Elderly				
Adult (15 - 64 years) vs. Elderly				
Old (25-65+years) vs. Young (0-24 years)	1.770	1.09-2.86	5.3454	0.0208
Sex at Birth				
Male vs. Female	1.060	0.75-1.49	0.106	0.7447
Race/Ethnicity (White = reference group)				
Black vs. White	2.060	1.29-3.28	51.58	<0.0001
Hispanic/Latino vs. White	0.329	0.18-0.61	18.67	<0.0001
Other vs. White	0.550	0.28-1.09	2.41	0.121
Origin				
US-born vs. Foreign-born	5.570	3.66-8.47	64.47	<0.0001

(table continues)

HIV Status				
Positive vs. Negative	1.390	0.78-2.50	1.2322	0.267
Homelessness				
Yes vs. No	2.010	0.99-4.06	3.8052	0.0511
Correctional Institution				
Yes vs. No	1.790	0.56-5.67	0.9666	0.3255
Alcohol				
Yes vs. No	2.030	1.40-2.94	13.93	0.0002
Illicit Drug Use				
Yes vs. No	2.060	1.28-3.32	8.7344	0.0031
Non-Injection Drug Use				
Yes vs. No	2.150	1.32-3.50	9.5823	0.002
Injection Drug Use				
Yes vs. No	0.799	0.21-3.01	0.1101	0.74
Substance Abuse				
Yes vs. No	2.260	1.30-3.93	8.26	0.004
Previous TB Diagnosis				
Yes vs. No	2.320	0.92-5.83	3.192	0.074

Statistically significant at alpha <0.05 bivariate association.
Significant at alpha <0.10, variable examined in the multivariable model.

Multivariable Analysis

Additionally, the second research question asks *what are the risk factors of genotype clustering among incident South Carolina TB cases from 2005 to 2011, when controlling for significant covariates?* To answer this question a stepwise multivariable analysis was performed. The stepwise multiple logistic regression model included all covariates that were significant in the bivariate modeling with an alpha of <0.10. These were origin, race, age, substance abuse, homelessness, and previous TB diagnosis. HIV status, correctional institution, and sex were not included in this stepwise analysis because they were not significant in the bivariate analysis. The results of the stepwise

multivariable analysis are shown in Tables 12 and 13 below (the full stepwise selection process is shown in Appendix B, Table B1).

Table 12

Odds ratios from the Best-Fit Logistic Regression Analyses of Tuberculosis Genotype Clustering and Associated Risk Factors

N=627, Clustered (n=382), Not Clustered (n=245)

Main Effect	Odds Ratio Estimates (95% C.I.)[a]	Wald X^2	*Wald p-value*
Origin, *US-Born vs. Foreign-Born*	5.67 (2.12% to 15.14%)	11.998	0.0285
Race, *Black vs. White*	1.96 (1.26% to 3.14%)	9.060	0.0005

[a]Wald 95% confidence intervals given in parentheses. Only significant predictors are listed.

Table 13

Model Fit Statistics from the Best-Fit Logistic Regression Analyses of Tuberculosis

	Testing Global Null Hypothesis: BETA=0		
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	71.574	1	<.0001
Score	72.485	1	<.0001
Wald	64.465	1	<.0001

Model Convergence Status: Convergence criterion (GCONV=1E-8) satisfied.

In the multivariable analysis, the only factors that remained significantly predictive for genotype clustering were being U.S. born, and being of Black race. Specifically, people US-born were over 5.5 times more likely to be part of a genotype cluster than those of foreign birth. People of Black race were almost 2 times more likely to be part of a genotype cluster compared to people of White race. As shown in Table 12, this final model was a good fit for genotype clustering with a large and significant Likelihood Ratio and Wald Score.

As part of the multivariable stage of the analysis, interaction terms were also tested. The interaction terms were tested in separate versions of the multivariable model where some insignificant covariates were added in order to assess the possibility of interaction. The most common method of incorporating an interaction term in a multivariable model is to create a product term. I did this by creating a variable whose value is the product of two independent variables (i.e., the 2 variables for which I was assessing a possible interaction multiplied by each other). Because a product term describes the relationship between two risk factors and an outcome, it can only be interpreted as an interaction if the two risk factors are in the model. Interaction terms were chosen because both of the covariates were significant in the crude model, and/or previous research has observed effect modification. The interaction terms that were tested were (1) origin and race, (2) origin and previous TB disease, (3) origin and substance abuse, (4) origin and HIV status, (5) previous TB disease and race, (6) substance abuse and race, and (7) HIV status and race. None of these interaction terms was significant in an adjusted model, and they were not included in the final model. Please see Table 11 above for the final model. For reference, Appendix B Table B2 shows the tested models including interaction terms.

Summary

This was a non-experimental, cross-sectional population-based molecular epidemiological study of TB in South Carolina from 2005 to 2011. The purpose of the study was to estimate the rate of recent TB transmission and investigate the relationship between TB genotype clustering and potential risk factors. The data was analyzed using

the "*n - 1 method"* of calculating recent TB transmission and multiple logistic regression to examine risk factors for clustering. The results indicated that about 50% of TB in South Carolina might be due to recent transmission. Additionally, the study revealed U.S. birth and Black race were independently and significantly associated with being part of a TB genotype cluster of two or more cases. In Chapter 5, I will further discuss the results of the research questions including interpretations, limitations of the study, implications for social change, and recommendations for future research.

Chapter 5: Conclusions and Recommendations

The purpose of this research was to estimate the proportion of South Carolina TB that may be due to recently acquired infection, and to determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates. In doing so, my objective was to provide suggestions to the South Carolina TB control program informed by the molecular epidemiology of TB in the state, as well as to articulate recommendations for future research. This was a non-experimental, cross-sectional, population-based molecular epidemiological study of TB in South Carolina from 2005 to 2011. The study population included 685 genotyped TB cases. I answered RQ1 using the *"n - 1 method"* for TB recent transmission estimation, and RQ2 using multiple logistic regression analysis to examine the association between TB clustering and predictors. The study results strongly suggested that about 50% of TB in South Carolina has been recently transmitted. Further, the analysis revealed several associations. The results of the bivariate analysis indicated a positive relationship between being U.S. born and genotype clustering, being Black and genotype clustering, older age and genotype clustering, being an alcoholic and genotype clustering, being a drug abuser and genotype clustering, and being both an alcohol and drug abuser and genotype clustering. The bivariate analysis also indicated that being Hispanic was protective for genotype clustering. In the multivariable analysis, the only factors that remained significantly predictive for genotype clustering were being U.S. born and being Black. In the next section, I interpret these findings.

Interpretation of Findings

Before discussing the research questions individually, I will discuss the results of the descriptive statistics performed. The study population (685 genotyped cases) and the entire state case population (1,346) were very similar with regard to the covariates of interest (see Table 8). Both were mostly adults, and the majority were men. Black was the most prevalent race, followed by White, and then by Hispanic and other. For both populations, close to 80% of the cases were born in the U.S. Both populations contained less than 9% HIV positive, less than 7% homeless, and less than 3% were incarcerated at the time of TB diagnosis. About 30% of South Carolina TB cases included in the study reported alcohol abuse, compared to about 20% for all TB cases from 2005-2011. The rates of drug abuse, particularly injection drug abuse, were low among both populations. For over 95% of TB cases in South Carolina this was their first TB diagnosis and this was also true for the genotyped cases (i.e., the study population). Chi-square analysis indicated that the study population was a reasonable representation of the entire case population of the state with the exceptions of that genotyped cases were older, more often alcoholic, and more often drug abusers (see Table 8). Particularly, children were underrepresented in the study sample ($X^2 = 46.4$, *p-value* < 0.0001).

South Carolina's estimated rate of recent transmission of about 50% is in keeping with previous studies in the southeast U.S., if somewhat higher. Kempf's 2005 Alabama study observed a recent transmission rate of 35% (Kempf, 2005). Talarico's 2011 Arkansas study did not estimate recent transmission but observed a high clustering rate of about 39% (Talarico, 2011). While Weis's 2002 Tarrant County, Texas, study

estimated a rather high recent transmission rate of about 36%, and Moonan's US-wide study estimated a recent transmission rate of about 23% (Weis, 2002, Moonan, 2012). In comparison with these previous findings, this study's estimate of recent transmission is higher. The estimate of recent transmission of 50% is outside the range of previous studies for low incidence countries with the highest estimate noted from a Southern California study of 45% (Rodwell, 2012). If this study's estimate of 50.2% for recent transmission suffers from bias, it is likely biased towards clustering (indicating greater recent transmission). This is because the study sample was older, and had more substance abuse than the general TB case population. Further, anecdotally, if there was any non-random influence for reasons why certain cases might be sent in for genotyping (and others not) it could be because contact tracers actually already suspected the case was part of an outbreak or cluster. The extent to which this may have resulted in sampling bias is unknown. However, even given these assertions, I would conclude that 50% might be an overestimate of recent transmission, there is good analytical reason to believe South Carolina's rate of recent transmission is very high. If I calculate a range of recent transmission based on the entire denominator of 1,346 confirmed cases we would still have a recent transmission rate of 25.5 % (343/1346) at the lowest; and this would be assuming no other cases in the state were clustered which is a rather far-fetched assumption. In reality, the recent transmission rate likely lies somewhere between 25.5 to 50.2%. This means that a staggering 1 to 2 out of every four cases of TB in South Carolina are due to recent transmission. This is quite substantial for a wealthy First World country with a low incidence of TB overall. Some European and

U.S. studies observed similar findings. An English study from 1998 showed a 52% rate of clustering and 40% rate of recent transmission; a national study carried out in the Netherlands during 1993–1997 found estimates for clustering and recent transmission of 46% and 35%, respectively; and a study of 7 sentinel surveillance sites in the USA during 1996–2000 estimated clustering at 48%; and a cross-sectional study during 2005–2009 estimated recent transmission as 23%, (Love, 2009; Moonan, 2012; Ellis, 2002; van Soolingen, 1999).

With regard to risk factors for clustering, the results of this study are fairly consistent with what has been observed in prior U.S. studies. Almost all U.S. studies have found U.S. birth to be an independent predictor for clustering. US-born persons would have had more time in the U.S. to have acquired and spread TB to their contacts, than recently arrived immigrants. Additionally, it does make sense that infected persons born outside of the U.S. might be more likely to have acquired TB in their country of origin, as the U.S. has a relatively low incidence of TB compared to much of the developing world. Further, U.S. physicians may be more likely to recognize TB symptoms and test for TB in immigrants from the developing world than in native U.S. persons. Thus, foreign-born persons may be more likely to have a singleton genotype if their TB is diagnosed early after arrival, whereas U.S. cases may go unrecognized longer allowing for more spread prior to intervention. To summarize foreign-born persons are more likely to have acquired their TB natively within a genotype cluster in their country of origin and may or may not propagate that cluster once they immigrate. Therefore, this finding is consistent and makes sense.

Almost all U.S. studies have observed Black race to be independently predictive of TB clustering and a few European studies have as well (Barnes, 1997; Cattamanchi, 2006; Ellis, 2006; Kempf, 2005; Cacho-Calvo, 2005). Importantly, 63% of the study population and 59% of South Carolina TB cases overall are of Black race, even though only about 30% of South Carolina's population is Black. Because TB disproportionately affects Black South Carolinians, this finding, while not surprising, is important. It would appear that in South Carolina, Black people are at much higher risk for TB transmission than other races. It is difficult to know if the risks Black people face are due to factors other than race. Because TB is a highly contagious, airborne disease for which close family contacts would be at risk, certainly, familial clusters will usually include a racial component as people that are biologically related to each other are likely to be of the same race. However, the literature indicates the higher risks experienced by Black people are almost certainly environmental (Kempf, 2005; Talarico, 2011; Weis, 2002). Recall this study's final model controlled for some environmental factors such as substance abuse, homelessness, origin, and a previous TB diagnosis; and Black race remained highly significant. Some hypotheses that have been presented in other studies include that southern Black Americans are adverse to the public health system because of past segregation, institutional racism, and an overall perception of mistrust (Mays, 2012; Mays, 2017; Wechkunanukul, 2016). This mistrust perception in people of Black race may result in being reluctant to seek diagnosis and treatment. Further when diagnosed, people of black race maybe less forthcoming to public health investigators regarding contacts, and well as contacts might be harder to

find and follow if they are attempting to avoid the government-based public health sector.

Unfortunately, this study was not able to access socioeconomic factors that may also play a role in the increased risk for people of Black race and TB clustering. Differences in the uninsured populations of Black and whites may certainly effect access to care and result in treatment and diagnosis delays causing higher rates of TB transmission, thus clustering. For Black people, the likelihood of being uninsured varies widely across states, however, uninsured rates for nonelderly Blacks are particularly high in the South. The uninsured rate for nonelderly Black people in South Carolina is between 23 and 30% (Duckett, 2013). In the US, Blacks are significantly more likely than Whites to be uninsured with more than 1 in 5 (21%) nonelderly Blacks uninsured, compared to 13% of their White counterparts (Duckett, 2013). Future research is needed to determine the reason or combination of reasons southern Black Americans are disproportionately affected by TB and how best to address these reasons.

Unlike my study, several U.S. studies have observed either alcohol, drug abuse, or all substance abuse as independently predictive for clustering. In an Arkansas, study (Talarico, 2011) rates of clustering were 19.8% among those with a history of alcohol abuse versus 8.6% among those without (*p-value* < 0.0001). Moonan's comprehensive US-wide study, noted earlier, also found a strong association between clustering and being more likely to abuse substances with an aOR of 1.4 (99% C.I. 1.3-1.7), where substance abuse was any abuse of drugs or alcohol (2012). In this study, substance abuse was significant in the bivariate model, and 33% of clustered cases were alcohol

abusers compared to only 20% of non-clustered cases. While 18% of clustered cases were drug, abusers compared to only 10% of non-clustered cases. As discussed earlier, my study sample was statistically different from the South Carolina case population with regard to alcohol and drug abuse. Alcoholic and drug addicts were overrepresented in the study sample, which may explain some of this perceived association; however, it is not likely to explain all of it. Importantly, recall from Chapter 2 that persons with substance abuse may be at greater risk for clustering for a number of reasons. They may (1) congregate closely in social settings such as bars (2) have poor recall of social contacts and events, thus are not able to provide a comprehensive contact list to public health investigators delaying or preventing PEP initiation, (2) have unknown, thus uncontrolled for co-morbidities that increase their vulnerability for disease as well as decrease the time they progress from infection to disease, or (3) be more prone to treatment non-adherence due to the interaction of alcohol and common TB drugs, (5) have poor nutrition and other substance abuse related health issues (6) persons who abuse alcohol, crack cocaine, heroin, marijuana, and methamphetamines have all been shown to experience significant weakening to one or more important immunologic mechanisms (Gamble, 2006; Baldwin, 1997; Lysle, 2000; Tashkin, 2002; Friedman, 2003; Mahajan, 2006) (7) substance abuse occurs in enclosed spaces with intentionally limited or poor ventilation and high volumes of human traffic likely increasing the odds of TB transmission (Oeltmann, 2006; Oeltmann, 2009), and (8) previous research has noted that persons who abuse substances are less likely to seek treatment resulting in an extended period of infectiousness and advanced disease at diagnosis (CDC, 2003). It

seems likely that contact tracing in these groups may not be as effective, leading to ongoing transmission.

In this study there were a few interesting findings that are somewhat inconsistent compared with prior research. First, current incarceration was not found to be associated with clustering. Recall that Kempf's Alabama study observed an aOR of 2.9 (95% C.I.1.3- 6.6) (2005) for the association of clustering and residence in a correctional facility *in the year prior to diagnosis*. However, in contrast, Moonan's U.S. wide population-based study did not observe an association with residence in a correctional facility at the time of diagnosis (aOR 0.8; 99% C.I. 0.7-1.0) (2012). Moonan's and this study likely had similar results (contrasting with Alabama's) regarding incarceration because our studies assessed only current incarceration, not prior or recent. This might suggest that the risk of TB transmission in incarceration settings may be more often due to short term stays in jails where TB often goes undiagnosed and untreated. Then the inmate is put back on the street to spread TB, where he or she might finally be diagnosed by the public health surveillance system.

Next, HIV status was not observed to be a significant predictor in my study as it has in a few prior U.S. studies (Ellis, 1994; Talarico, 2011; Moonan, 2012). This could be for a number of reasons. First, there were very few HIV positive individuals in my study population and this, coupled with the fact that there were many individuals in my study population that had to be dropped from analysis due to unknown HIV status (recall this was the primary variable which necessitated dropping cases), these low numbers substantially decreased my power in both the bivariate and multivariate

analysis to reveal an effect, if, in fact, one exists. Further, it is not inconceivable that people who refuse HIV testing do so because they have multiple HIV risk factors, or in fact, already know they are positive and do not share those results. This could have biased the findings. Also, this study had contained no information regarding immune-competence of any HIV positive individual. Thus, an HIV positive person that is receiving HIV therapy with adequate immune-competence may not be at a noticeably increased risk of being part of a TB cluster.

Limitations of Study

A few limitations should be noted that affect the interpretation of the findings and the generalizability of my results. First, my study was limited to those confirmed TB cases for which a valid TB genotype could be obtained. As discussed at length in Chapter 3, generally cases that are exclusively extrapulmonary (those are TB cases that have only extrapulmonary disease rather than both pulmonary and extrapulmonary) are confirmed via some other testing methodology such as a PCR instead of culture, thus there is not culture to genotype. This was true with the exception of extrapulmonary cases that are of the pleural space of the lungs, which are usually confirmed, similar to pulmonary cases using sputum specimen cultures. Another reason a case may have been considered ineligible for culture was the age of the patient. Patients, ages between 5 and 12 years, are difficult to obtain a positive culture from and cultures are often not attempted or used for confirmation. Once those confirmed cases considered "ineligible" for genotype were excluded only 1,080 of the 1,346 TB cases in South Carolina from 2005-2011 were eligible for genotype. Of these, 685 were successfully genotyped. The other cases were not

genotyped for a variety of reasons including a culture was not obtained because the patient was lost to follow-up, the patient had a negative culture but was confirmed via some other combination of tests or symptoms, or the diagnosing laboratory simply did not send their TB isolates to the National TB lab for genotyping. Thus, only 50.9% (685/1,346) of *all* TB confirmed cases were genotyped. The 49.1% of confirmed incident cases that were not genotyped may have differed from the 685 incident TB cases in important ways that may have biased my results. However, I was able to use publically available descriptive data on the TB incident case population of South Carolina from 2005 to 2011 to compare with my fully genotyped sample on all the covariates examined in my study. This information was available through an intranet query site called the CDC's Online Tuberculosis Information System (OTIS) that provides data in 5-year aggregate summary totals by state (CDC, 2016). From this, I was able to determine how alike or different my sample was from the entire sampling frame in at least a few characteristics (see Table 8). In addition, this limitation also represents an opportunity. There was a high proportion (43%) of TB cases without a culture, and therefore, not genotyped. This is an opportunity to emphasize the importance of comprehensive TB case genotyping, in order to fully characterize the molecular epidemiology of TB in any jurisdiction or geography. This is discussed further in the Recommendations section of this chapter.

This second limitation of this study was that it was limited in place and time, in that I am only able to examine genotypes of cases diagnosed within South Carolina between 2005 and 2011. Cases diagnosed outside of this time period (either before or after) within South Carolina, cases diagnosed within this time period but outside of

South Carolina, and cases diagnosed outside this time period (either before or after) and outside of South Carolina will not be included in this study. This means that TB chains of transmission beyond the geographic and time boundaries set in this study cannot be investigated. Certainly, this is a common limitation of all molecular epidemiological studies, as researchers will always be limited by the data they collect or, as in the case of this study, that which was secondarily available to them. However, because TB transmission recognizes no time or geographical boundaries, this will have implications for the conclusions that may be drawn from study's findings. These implications are explored in the Conclusions, Chapter 5, of this research.

A third limitation of this study was that for the purposes of answering Research Question 1(b) what proportion of TB in South Carolina may be due to recent transmission?. To answer this question epidemiological information (dates of onset, diagnosis, etc.) was not reviewed. This determination was a mathematical calculaton in order to provide a reasonable indication of the proportion of South Carolina TB cases that may be due to recent transmission. The epidemiological assertion is based on recent transmission index of *(RTI) i.e., RTIn-1 = (nc - c)/n*, in which *n* is the total number of the studied cases, *nc* is the total number of cases in cluster (size 2 or greater) and *c* is the number of genotypes represented by at least two cases. Based on this index, patients in cluster are considered as recent transmission and non-cluster cases are considered as reactivation. This index also has been referred to as the "*n - 1 method*." In depth review of source medical records to determine true source cases for comparison to the mathematically calculated determination were not the approach of this study for two

reasons. First, because the time it takes latent TB infection to manifest to disease varies widely as it was often a result of both known and unknown host factors, dates of onset and diagnosis obtained may not be helpful or reliable in determing source or index cases for clusters or outbreaks. Second, most TB population-based molecular epidemiological studies do not use medial record review to determine the proportion of recent transmission, but do it mathematically in just the way I have outlined in Research Question 1(b). This was an exceptible limitation because the methods used in this study were consistent with previous research making my research comparable to other studies of this nature.

Recommendations

The findings revealed that a considerable portion of TB in South Carolina was due to recent transmission, and U.S. birth and Black race were independently and significantly associated with being part of a TB genotype cluster of two or more cases. While incident cases of TB have decreased in South Carolina over the previous 8 years, going from 261 cases reported in 2005 to 140 cases reported (South Carolina DHEC, 2014), South Carolina still had the 15th highest case rate in the U.S. with 3 cases per 100,000 people (CDC, 2013). TB is a significant public health problem in South Carolina, it is with that in mind this study was undertook, and recommendations are made. While South Carolina has a relatively low incidence of TB, in general it does appear that most of their TB is due to recent transmission, not importation or reactivation; thus happening well within the purview of the United States' seemingly robust health care system. This

indicates that opportunities for prevention and interruption of the chain of transmission are being missed.

Based on the process and results of this study, I would make several specific recommendations to the South Carolina TB Control Program.

1. From 2005-2011, 37% of eligible TB cases in South Carolina did not have a genotyped culture and 49.1% of all TB cases did not have a genotyped culture. The proportion of cases in South Carolina being genotyped is too low. While it is difficult to obtain a culture on all TB cases for a variety of reasons (reasons discussed at length in earlier chapters), every effort should be made at the local and state level to make this happen. Every TB case should be confirmed with a culture and every culture should be genotyped, especially children. Understandably, children are difficult on which to obtain cultures, but they are also the most important. Because children progress quickly for TB infection to TB disease, they might be sentinel cases in a TB cluster or outbreak. Further, TB in children is usually an indication of new and ongoing transmission. Confirming pediatric cases with a culture and knowing their genotypes early is of the utmost importance. Health department staff should follow-up with private sector providers and labs to make sure cultures are sent in for genotyping. There are important reasons to obtain a genotype on every case that go well beyond having complete research data. They include:
    a. identification or confirmation of outbreaks,
    b. identification of locations at which transmission occurs,
    c. identification of characteristics of TB disseminators,

d. characteristics of people at increased risk for acquiring TB infection and for rapid progression to TB,
e. detection of unsuspected transmission,
f. determination of the potential for reinfection,
g. detection of laboratory cross-contamination and
h. evaluation of TB control activities designed to prevent transmission.

2. The TB case record and accompanying RVCT is not a static document, it should be continuously updated as new information is revealed. While the issue of missing data was not a problem in this study, it was apparent in data review that some patient characteristics were much less complete than those examined by this study. For instance, patient employer/occupation contained many unknown/missing data. In addition, the question of whether antibiotic susceptibilities were performed, and the associated results of those susceptibilities also contained a lot of unknown/missing information. Having more complete case records is important for conducting formal studies such as this one, as well as evaluating the effectiveness of prevention and control efforts.
3. In South Carolina, measures should be implemented to address the staggering racial disparity in TB disease. This will be daunting task, but should begin with a few initial steps. First, the racial and ethnic makeup of the DHEC TB control staff should better reflect the TB case population. Additionally, language, ethnic, cultural, and economic barriers should not be obstacles to prevent thorough case and contact investigations. Community partners may be enlisted such as churches and

community health centers. Where there have been other areas of public health success in addressing racial disparities such as diabetes education, smoking cessation promotion, and HIV control and prevention, determine if there are opportunities for TB education to piggyback off some of these efforts.

4. While substance abuse was not significant in the final model, this study revealed that substance abusers were clustered more than not. Alcohol and drug environments are likely places where TB is being spread. Drinking and drugging acquaintances should not be overlooked in contact investigations. Understandably, a cases' addiction behavior is a sensitive topic of discussion however, contact tracers should repeatedly query it.
5. While recent homelessness was only a significant predictor in the crude model, homeless shelters or camps might also be places where TB is being spread. Street friends, "couch surfing" contacts, and other transient living environments should be investigated thoroughly. Homeless adults and unemployed adults stating they have estranged family or no family should not be assumed to have no contacts.
6. HIV status was not a significant predictor of TB clustering in this study population. However, HIV status was unknown 8% of these cases. It is important that every TB case be offered and strongly encouraged to take an HIV test. Further, contacts to infectious TB cases should be tested for HIV in order to prioritize them for treatment.
7. Contact investigations must move beyond the traditional approach. Typically, the South Carolina TB Control Program uses the concentric circle approach to prioritize

contact investigations. Several studies have shown the concentric circle method often misses peripheral, yet high risk contacts (Moro, 2002, Izumi, 2015, Feske, 2011, Driver, 2006). This is probably resulting in contacts and cases being missed. Alternatives to the traditional contact investigation should be piloted in South Carolina. This could include social network contact tracing, or activity-space-contact-tracing.

8. "Lost to follow-up" status is unacceptable for TB cases. Because TB is a highly infectious and, when untreated, a sometimes-fatal disease, all efforts should be made to avoid losing any case for follow-up and treatment.

The current study reveals several opportunities for future research. A southeast U.S. study with better representation among children is needed. More research is needed to flesh out the relationship between race and genotype clustering. Specifically, studies need to be conducted to identify the root causes for the racial disparities in TB disease. More research is needed to better understand if there is a relationship between being HIV positive and being part of a TB cluster. A study with complete information on co-morbidities such as diabetes, COPD, and smoking may help present a more complete model for risk of TB clustering as certain co-morbidities that increase the risk of disease progression. Studies or pilot programs investigating alternatives to the traditional contact investigation should be implemented. A South Carolina study similar to a Tokyo study by Izumi and colleagues whose aim was to identify possible tuberculosis hotspots using TB genotype clustering statuses and the concept of "activity space", which is a place where patients spend most of their waking hours might be

very informative for determining where to target control efforts (Izumi, 2015). Also, a study using Social Network Analysis might be indicated.

Implications for Social Change

As stated previously, the goal of the research was to estimate the proportion of South Carolina TB that may be due to recently acquired infection, and determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates. The research performed was able to answer the research questions and provide insight into the previously unknown transmission dynamics and molecular epidemiology of TB in South Carolina. The study findings provide opportunities for social change. The recommendation section above provides useful information for TB services and policy makers to help identify where resources may be best deployed.

The relationships revealed indicate more could be done to interrupt the TB chain of transmission at the individual and local level. At the individual level as part of the positive social change implications of this research it is my hope that cases that would have been missed when investigated with traditional contact investigation alone, will be discovered when employing targeted contact investigation informed by the findings of this study. At the organizational level, genotyping may facilitate quicker confirmation of known contacts, detection of unknown contacts, or revelation of transmission environments (Malakmadze, 2006; Yeo, 2006). If a jurisdiction participates in the U.S. national TB genotyping program, consistently sends all their confirmed case specimens in for genotyping, and checks those results frequently they may find linked cases that their standard contact investigations did not reveal. In addition, TB contact tracers may

confirm suspected epidemiological links they were already investigating. This approach may help a jurisdiction make decisions about how aggressive and widespread their contact investigations should or should not be, where to focus limited resources, and if there are transmission environments traditional contact investigations are missing. Targeted contact-investigation informed by TB genotype results could result in an overall cost-savings to the South Carolina TB Control Program and similar programs.

At the societal level, the social change implication is that it might facilitate reaching the U.S. TB elimination goals by informing the South Carolina TB Control Program and potentially other states' control efforts, particularly states with similar populations and public health resources as that of South Carolina. As discussed in the recommendations section above, the study's findings provide a platform for future research. This research could lead to better TB control efforts that result in an interruption in the chain of transmission and resulting in fewer TB infections. This would bring the U.S. one-step closer to TB elimination.

Conclusion

The purpose of this research was to estimate the proportion of South Carolina TB that may be due to recently acquired infection, and determine the risk factors associated with the genetic clustering of identical *M. tuberculosis* isolates. The results indicated that a considerable portion of TB in South Carolina was due to recent transmission. Additionally, the results showed U.S. birth and being of Black race were independently and significantly associated with being part of a TB genotype cluster of two or more cases. The key messages of this study are (a) a substantial portion of TB in South

Carolina is due to recent transmission not reactivation or importation, and (b) transmission of TB in South Carolina occurs in groups often defined by American origin, and Black race. These important points reveal one important conclusion; most TB in South Carolina is preventable. Because the global and domestic impact of TB disease is substantial with global incidence by recent estimates nearly 9.6 million people, and in the U.S. between 500 and 600 people dying every year from TB (WHO, 2016), TB elimination is an important public health goal. TB although an ancient plague is now preventable, treatable, and curable. TB elimination in the developed world is within reach.

References

Asiimwe, B, Joloba, M, Solomon, G, Koivula, T, Kateete, D, Katabazi, F, Pennhag, A, Petersson, R, & Kallenius, G (2009). DNA restriction fragment length polymorphism analysis of mycobacterium tuberculosis isolates from HIV-seropositive and HIV seronegative patients in Kampala, Uganda. *British Medical Infectious Diseases*, (9)12, 1-9.

Baldwin, G, Tashkin, D, Buckley, D, Park, A, Dubinett, S, & Roth, M. (1997). Marijuana and cocaine impair alveolar macrophage function and cytokine production. *American Journal of Respiratory Critical Care Medicine*, 156(5), 1606-1613.

Barnes, P, & Cave, M. (2003). Molecular epidemiology of tuberculosis. *New England Journal of Medicine*, 349(12), 1149-1156.

Barnes, P, Yang, Z, Preston-Martin, S, Pogoda, J, Jones, B, Otaya, M, Eisenach, K, Knowles L, Harvey, S, & Cave, M. (1997). Patterns of tuberculosis transmission in central Los Angeles. *Journal of the American Medical Association*, 278(14), 1159-1163.

Bass, P, Farer L, Hopewell, P, Jacobs, R, & Snider, D. (1990). Diagnostic standards and classification of tuberculosis. *American Review of Respiratory Diseases*, 142(1), 725-735.

Bidovec-Stojkovič, U, Seme, K, Žolnir-Dovč, M, & Supply, P. (2014). Prospective genotyping of *mycobacterium tuberculosis* from fresh clinical samples. *Public Library of Science One*, 9(10), e109547.

Borgdorff, M, Sebek, M, Geskus, R, Kremer, K, Kalisvaart, N, & van Soolingen, D.

(2011). The incubation period distribution of tuberculosis estimated with a molecular epidemiological approach. *International Journal of Epidemiology*, 40(4), 964-970.

Cacho Calvo, J, Astray Mochales, J, Pérez Meixeira, A, Ramos Martos, A, Hernando García, M, Sánchez Concheiro, M, Domínguez Pérez, J, Gómez, A, Samper, S, & Martín, C. (2005). Ten-year population-based molecular epidemiological study of tuberculosis in the metropolitan area of Madrid, Spain. *International Journal of Tuberculosis and Lung Diseases*, 9(11), 1236-1241.

Cantwell, M, McKenna, M, McCray, E, & Onorato, I. (1998). Tuberculosis and race/ethnicity in the United States: impact of socioeconomic status. *American Journal of Respiratory Critical Care Medicine*, 157(4 Pt 1), 1016-1020.

Cattamanchi, A, Hopewell, P, Gonzalez, L, Osmond, D, Masae Kawamura, L, Daley, C, & Jasmer, R. (2006). A 13-year molecular epidemiological analysis of tuberculosis in San Francisco. *International Journal of Tuberculosis and Lung Disease*, 10(3), 297-304.

Centers for Disease Control and Prevention, Division of Tuberculosis Elimination. (2009). CDC Tuberculosis Surveillance Data Training Report of Verified Case of Tuberculosis (RVCT) Instruction Manual. Published June 2009.

Centers for Disease Control and Prevention, Division of Tuberculosis Elimination. (2011). *Core Curriculum on Tuberculosis: What the Clinician Should Know, 5th edition, 21-24.*

Centers for Disease Control and Prevention. (2016). Online Tuberculosis Information

System.

http://wonder.cdc.gov/TB-v2014.html. Assessed on August 13, 2016.

Centers for Disease Control and Prevention. (2006). Racial/ethnic disparities in diagnoses of HIV/AIDS --- 33 States, 2001-2004. *Morbidity and Mortality Weekly Report*, 55(5), 121-125.

Centers for Disease Control and Prevention. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, and Division of Tuberculosis Elimination (DTBE). (2013). *Reported Tuberculosis in the United States*, 2012.

Centers for Disease Control and Prevention. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, and Division of Tuberculosis Elimination (DTBE). (2014). *Reported Tuberculosis in the United States*, 2014.

Center for Disease Control, (2012). South Carolina-2010 Profile. http://www.cdc.gov/nchhstp/stateprofiles/pdf/South_Carolina_profile.pdf . Retrieved May 1, 2014.

Centers for Disease Control (2011). Surveillance of health Status in minority communities --- racial and ethnic approaches to community health across the U.S. (REACH U.S.) Risk Factor Survey, United States, 2009. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 60(SS06), 1-41.

Centers for Disease Control (2003). Transmission of *mycobacterium tuberculosis* associated with failed completion of treatment for latent tuberculosis infection Chickasaw County, Mississippi, June 1999–March 2002. *Morbidity and Mortality Weekly Report*, 52, 222-224.

Centers for Disease Control (2003). Tuberculosis outbreak among homeless persons—King County, Washington, 2002–2003. *Morbidity and Mortality Weekly Report*, 52, 1209–1210.

Centers for Disease Control (2005). Tuberculosis transmission in a homeless shelter population—New York, 2002–2003. *Morbidity and Mortality Weekly Report*, 54, 149–151.

Chan-Yeung, M, Kam, K, Leung, C, Wang, J, Yew, W, Lam, C, & Tam, C. (2006). Population-based prospective molecular and conventional epidemiological study of tuberculosis in Hong Kong. *Respirology,* 11(4), 442-448.

Chin, D, Crane, C, Diul, M, Sun, S, Agraz, R, & Taylor, S. (2000). Spread of *mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. *Journal of the American Medical Association*, 283(22), 2968- 2974.

Comas, I, & Gagneux, S. (2009). The past and future of tuberculosis research. *Public Library of Science Pathogens*, 5(10), e600-607.

Daley, C, Small, P, & Schecter, G. (1992). An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus: an analysis using restriction-fragment-length polymorphisms. *New England Journal of Medicine*, 326(4), 231-235.

Department of Health and Human Services (DHHS). (2013). Centers for Disease Control and Prevention, Justification of Estimates for Appropriations Committees. Retrieved from www. cdc.gov.

Diel, R, Schneider, S, Meywald-Walter, K, Ruf, C, Rüsch-Gerdes, S, & Niemann, S. (2002). Epidemiology of tuberculosis in Hamburg, Germany: long-term population-based analysis applying classical and molecular epidemiological techniques. *Journal of Clinical Microbiology*, 40(2), 532-539.

Di Perri, G, Cruciani, M, & Danzi, M. (1989). Nosocomial epidemic of active tuberculosis among HIV-infected patients. *Lancet*, 2(8678-8679), 1502-1504.

Driver, C, Kreiwirth, B, Macaraig, M, Clark, C, Munsiff, S, Driscoll, J, & Zhao, B. (2006). Molecular epidemiology of tuberculosis after declining incidence, New York City, 2001-2003. *Epidemiology and Infection*, 135, 634-643.

Driver, C, Macaraig, M, McElroy, P, Clark, C, Munsiff, S, Kreiswirth, B, Driscoll, J, & Zhao, B. (2006). Which patients' factors predict the rate of growth of *mycobacterium tuberculosis* cluster in an urban community*?*. *American Journal of Epidemiology*, 161(1), 21-31.

Duckett, P, & Artiga, S. (2013). Health Coverage for the Black Population Today and Under the Affordable Care Act. *Disparties Policy,* July 23, 2013.

Ellis, B, Crawford, J, Braden, C, Scott, J, McNabb, S, Moore, M, Kammerer, S, & National Tuberculosis Genotyping and Surveillance Network Work Group. (2002). Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerging Infectious Disease*, 8(11), 1197-1209.

Fendrich, M, Johnson, T, Wislar, J, Hubbell, A, & Spiehler, V. (2004). The utility of drug testing in epidemiological research: results from a general population survey. *Addiction*, 99(2), 197-208.

Feske, M, Teeter, L, Musser, J, & Graviss, E. (2011). Giving TB wheels: Public transportation as a risk factor for tuberculosis transmission. *Tuberculosis*, 91, S16eS23.

Fitzpatrick, L, Hardacker, J, Heirendt, Agerton, T, Streicher, A, Melnyk, H, Ridzon, R, Valway, S, & Onorato, I. (2001). A preventable outbreak of tuberculosis investigated through an intricate social network. *Clinical Infectious Disease*, 33(11), 1801-1806.

Fok, A, Numata ,Y, Schulzer, & FitzGerald, M. (2008). Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. *International Journal of Tuberculosis and Lung Disease*, 12(5), 480-492.

Franzetti, F, Condecase, L, Mattelli, A, Esposti, A, Bandera, A, Lacchini, C, Lombardi,A, Pinsi, G, Zanini, F, El-Hamad, I, & Gori, A. (2010). Genotyping analyses of tuberculosis transmission among immigrant residents in Italy. *Clinical Microbiology and Infection*, 16(8), 1149-1154.

Friedman, H, Newton, C, & Klein, T. (2003). Microbial infections, immunomodulation, and drugs of abuse. *Clinical Microbiology Review*, 16(2), 209-219.

Gamble, L, Mason, C, & Nelson, S. (2006). The effects of alcohol on immunity and bacterial infection in the lung. *Médecine et Maladies Infectieuses*, 36(2), 72-77.

Hamblion, E, Le Menach, A, Anderson, L, Lalor, M, Brown, T, Abubakar, A, Anderson, C Maguire, H & Anderson, S (2016). Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from

first 3 years of universal MIRU-VNTR strain typing. *Thorax*, 71(8), 749–756.

Hermans, P, van Soolingen, D, Dale, J, Schuitema, A, McAdam, R, Catty, D, & van Embden, J. (1990). Insertion element IS986 from *mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *Journal of Clinical Mirocrobiology*, 28(9), 2051-2058.

Hernandez-Garduno, E, Kunimoto, D, Wang, L, Rodrigues, M, Elwood, K, Black, W, Mak, S, & FitzGerald, J. (2002). Predictors of clustering of tuberculosis in greater Vancouver: a molecular epidemiologic study. *Canandian Medical Association Journal*, 167(4), 349-352.

Hill, B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295-300.

Hosmer, D, & Lemeshow, S. (1989). *Applied Logistic Regression.* New York, NY: John Wiley & Sons.

Hopewell, P, & Bloom, B. (2000). Tuberculosis and other mycobacterial diseases. In: Murray, J, Nadel, J. eds. *Respiratory Medicine* (3rd ed., 1043-1105). Philadelphia, PA, WB Saunders Company,

Izumi, K, Ohkado, A, Uchimura, K, Murase, Y, Tatsumi, Y, Kayebeta, A, Watanabe, Y, & Ishikawa, N (2015). Detection of tuberculosis infection hotspots using activity spaces based spatial approach in urban Tokyo, from 2003 to 2011. *Public Library of Science One*, 10 (9).

Jonsson, J, Hoffner, S, Berggren, I, Bruchfeld, J, Ghebremichael, S, Pennhag, A & Groenheit, R. (2014). Comparison between RFLP and MIRU-VNTR genotyping

of *Mycobacterium tuberculosis* strains isolated in Stockholm 2009 to 2011. *Public Library of Science One*, (10) 13.

Kamper-Jørgensen, Z, Andersen, A, Kok-Jensen, A, Bygbjerg, I, Andersen, P, Thomsen V, Kamper-Jørgensen, M, & Lillebaek, T. (2012). Clustered tuberculosis in a low-burden country: nationwide genotyping through 15 years. *Journal Clinical Microbiology*, 50(8), 2660-2667.

Kempf, M, Dunlap, N, Lok, K, Benjamin Jr, W, Keenan, N, & Kimerling, M. (2005). Long-term molecular analysis of tuberculosis strains in Alabama, a state characterized by a largely indigenous, low-risk population. *Journal of Clinical Microbiology*, 43(2), 870-878.

Kirenga, B, Ssengooba, B, Muwonge, C Nakiyingi, L, Kyaligonza, S, Kasozi, S, Mugabe, F, Boeree, M, Moses, & Okwera, A. (2015). Tuberculosis risk factors among tuberculosis patients in Kampala, Uganda: implications for tuberculosis control. *BioMed Central Public Health*, 15(13).

Kleinbaum, D, Kupper, L, & Muller, K (1988). Applied Regression Analysis and Other Multivariable Methods. Boston, PWSKent.

Kulaga, S, Behr, M, Musana, K, Brinkman, J, Menzies, D, Brassard, P, Kunimoto, D, Tannenbaum, T, Thibert, L, Joseph, L, Boivin, J, & Schwartzman, K. (2002). Molecular epidemiology of tuberculosis in Montreal. *Canadian Medical Association Journal,* 167(4), 353-354.

Lee, J. (2015). Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberculosis Respiratory Disease (Seoul)*, 78(2), 47-55.

Leonhardt, K, Gentile, F, Gilbert, B, & Aiken, M. (1994). A cluster of tuberculosis among crack house contacts in San Mateo County, California. *American Journal of Public Health*, 84(11), 1834-1836.

Love, J Sonneneber, P, Glynn, J, Gibson, A, Gopaul, K, Fang, Z, & Le Brun, F. (2009). Molecular epidemiology of tuberculosis in England, 1998. *International Journal of Tuberculosis and Lung Disease*, 12(2), 201-207.

Lysle, D, & How, T. (2000). Heroin modulates the expression of inducible nitric oxide synthase. *Immunopharmacology*, 46(3), 181-192.

Mays, V. (2012). The legacy of the U. S. Public Health Services study of untreated syphilis in African American men at tuskegee on the Affordable Care Act and Health Care Reform fifteen years after President Clinton's apology. *Ethics Behavior*, 22(6):411-418.

Mays, V, Jones, A, Delany-Brumsey, A, Coles, C, & Cochran, S. (2017). Perceived discrimination in health care and mental health/substance abuse treatment among Blacks, Latinos, and Whites. *Medical Care*, 55(2):173-181.

Maguire, H, Dale, J, McHugh, T, Butcher, P, Gillespie, S, Costetsos A, Al-Ghusein, H, Holland, R, Dickens, A, Marston, L, Wilson, P, Pitman, R, Strachan, D, Drobniewski, F, & Banerjee, D. (2003). Molecular epidemiology of tuberculosis in London 1995-7 showing low rate of activee transmission. *Journal of Clinical Pathology: Molecular Pathology*, 56(7), 121-126.

Mahajan, S, Hu, Z, Reynolds, J, Aalinkeel, R, Schwartz, & S, Nair, M (2006). Methamphetamine modulates gene expression patterns in monocyte derived

mature dendritic cells: implications for HIV-1 pathogenesis. *Molecular Diagnosis & Therapy,* 10(4), 257-269.

Malakmadze, N, González, I, Oemig, T, Isiadinso, I, Rembert, D, McCauley, M, Wand, P, Diem, L, Cowan, L, Palumbo, G, Fraser, M, & Ijaz, K. (2005). Unsuspected recent transmission of tuberculosis among high-risk groups: implications of universal tuberculosis genotyping in its detection. *Clinical Infectious Disease*, 40(3), 366-373.

McAdam, J, Bucher, S, Brickner, P, & Lascher, S. (2009). Latent tuberculosis and active tuberculosis disease rates among the homeless, New York, New York, USA, 1992-2006. *Emerging Infectious Diseases*, 15(7), 1109-1111.

McElroy, P, Southwick, K, Fortenberry, E, Levine, E, Diem, L, Woodley, C, Williams, P, McCarthy, K, Ridzon, R, & Leone, P. (2003). Outbreak of tuberculosis among homeless persons coinfected with human immunodeficiency virus. *Clinical Infectious Diseases*, 36(10), 1305-1312.

Mensch, B, & Kandel, D. (1988). Underreporting of substance use in a national longitudinal youth cohort: individual and interviewer effects. *Public Opinion Quarterly*, 52(1), 100-124.

Midanik, L. (1988). The validity of self-reported alcohol use: a literature review and assessment. *British Journal of Addiction*. (83), 1019-29.

Miller, A, Sharnprapai, S, Suruki, R, Cickren, E, Nardell, E, Driscoll, J, McGarry M, Taber, H, & Etkind, S. (2002). Impact of genotyping of *Mycobacterium tuberculosis* on public health practice in Massachusetts. *Emerging Infectious*

Diseases, 8(11), 1285-1289.

Moonan, P, Ghosh, S, Oeltmann, J, Kammerer, J, Cowan, L, & Navin, T. (2012). Using genotyping and geospatial scanning to estimate recent *Mycobacterium tuberculosis* transmission, United States. *Emerging Infectious Diseases*, 18(3), 458–465.

Moro, M, Salamina, G, Gori, A, Penati, V, Sacchetti, R, Mezzetti, F, Infuso, A, & Sodano, L. (2002). Two-year population-based molecular epidemiological study of tuberculosis transmission in the metropolitan area of Milan, Italy. *European Journal of Clinical Microbiology and Infectious Disease*, 21(2), 114-122.

Oeltmann, J, Kammerer, J, Pevzner, E, & Moonan, P. (2009). Tuberculosis and substance abuse in the United States, 1997–2006. *Archives of Internal Medicine*, 169, 189–197.

Oeltmann, J, Oren, E, Haddad, M, Lake, L, Harrington, T, Ijaz, K, & Narita, M. (2006). Tuberculosis outbreak in marijuana users, Seattle, Washington, 2004. *Emerging Infectious Diseases,* 12(7), 1156-1159.

Online Tuberculosis Information System (OTIS), National Tuberculosis Surveillance System, United States, 1993-2013. U.S. Department of Health and Human Services (U.S. DHHS), Centers for Disease Control and Prevention (CDC), Division of TB Elimination, CDC WONDER Online Database, April 2015. Retrieved from http://wonder.cdc.gov/tb-v2013.html

Reif, S, Whetten, K, & Wilson, E. (2012). HIV/AIDS Epidemic in the South Reaches Crisis Proportions in Last Decade [White paper]. Duke Center for Health Policy and Inequalities Research. Retrieved from http://www.hivlawandpolicy.org/

Reza Allahyar Torkaman, M Nasiri, M, Farnia, P, Hasan Shahhosseiny, M Mohadeseh, M & Velayati, A. (2014). Estimation of recent transmission of *mycobacterium tuberculosis* strains among Iranian and Afghan immigrants: A cluster-based study. *Journal of Clinical and Diagnostic Research,* 8(9), DC05–DC08.

Ricks, P, Cain, P, Oeltmann, E, Kammerer, S, & Moonan, P. (2011). Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the U.S., 2005-2009. *Public Library of Science One*, 6 (11), e27405.

Romelsjo, A, Leifman, H, & Nystrom, S. (1995). A comparative study of two methods for the measurement of alcohol consumption in the general population. *International Journal of Epidemiology,* 24, 929-36.

Rodwell, T, Kapasi, A, Barnes, R, & Moser, K. (2012). Factors associated with genotype clustering of *Mycobacterium tuberculosis* isolates in an ethnically diverse region of southern California, United States. *Infectious Genetic Evolution,* 12(8), 1917–1925.

Rothman, K, & Greenland, S. (2005). Causation and causal inference in epidemiology. *American Journal of Public Health,* 95(51) Suppl 1, S144-50.

Salinsky, E, (2010). Governmental public health: an overview of state and local public health agencies. *National Health Policy Forum*, background paper no. 77. Retrieved from www.nhph.com

Samper, S, Iglesias, M, Rabanaque, M, Lezcano, M, Vitoria, L, Rubio, M, Gómez-Lus, R, Gómez, L, Otal, I, & Martín, C. (1998). The molecular epidemiology of tuberculosis in Zaragoza, Spain: a retrospective epidemiological study in 1993. International *Journal of Tuberculosis and Lung Disease*, 2(4), 281-287.

Santic, Z, & Galic, C. (2013). Epidemiology of tuberculosis during the period 1703–2011: honoring the World Tuberculosis Day. *Materia Socio Medica*, 25(4), 291–294.

Serpa, J, Teeter, L, Musser, J, & Graviss, E. (2009). Tuberculosis disparity between US-born blacks and whites, Houston, Texas, USA. *Emerging Infectious Diseases*, 15(6), 899-904.

Small, P, Hopewell, P, Singh, S, Paz, A, Parsonnet, J, Ruston, D, Schecter, G, Daley, C, & Schoolnik, G. (1994). The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *New England Journal of Medicine,* 330(24), 1703-1709.

South Carolina Department of Health and Environmental Control (South Carolina DHEC), (2014). Reported tuberculosis cases in South Carolina by county 2004 – 2013. https://www.scdhec.gov/health/disease/tb/docs/South Carolina_TB.pdf .

Suffys, P, Ivens de Araujo, M, & Degrave, W. (1997). The changing face of the epidemiology of tuberculosis due to molecular strain typing - a review. *Memórias do Instituto Oswaldo Cruz*, 92(3), 297-316.

Tan de Bibiana, J, Rossi, C, Zwerling, A, Thibert, L, McIntosh, F, Behr, M, Menzies, D, & Schwartzman, K. (2011). Tuberculosis and homelessness in Montreal: a retrospective cohort study. *Biomed Central Public Health*, 11(833).

Talarico, S, Ijaz, K, Zhang, X, Mukasa, L, Zhang, L, Marrs, C, Cave, M, Bates, J, & Yang, Z. (2011). Identification of factors for tuberculosis transmission via an integrated multidisciplinary approach. *Tuberculosis,* 91(3), 244-249.

Tashkin, D, Baldwin, G, & Sarafian, T. (2002). Respiratory and immunologic

consequences of marijuana smoking. *Journal of Clinical Pharmacology*, 42 (11), ((suppl)) 71S- 81S.

Torgersen, J, Dorman, S. Baruch, N, Hooper, N, & Cronin, W. (2006). Molecular epidemiology of pleural and other extrapulmonary tuberculosis: a Maryland state review. *Clinical Infectious Disease*, 42(10),1375-1378.

Tsuchiya, T. (2006). Exogenous re-infection in tuberculosis. *Kekkaku*, 81(2), 79-91.

United States Department of State (U.S. DOS). (2009). Tuberculosis: global impact and U.S. response. Retrieved from http://2001-2009.state.gov/g/oes/rls/fs/2003/18798.htm

United States Government Global Tuberculosis Strategy 2015–2019: https://www.usaid.gov/sites/default/files/documents/1864/Reach-Cure-Prevent-2015-2019-TBStrategy.pdf .

van Soolingen, D, Borgdorff, M, de Haas, P, Sebek, M, Veen, J, Dessens, M, Kremer, K, & van Embden, J. (1999). Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *Journal of Infectious Disease*, 180(3) 726-736.

Verver, S, Warren, R, Munch, Z, Vynnycky, E, van Helden, P, Richardson, M, van der Spuy, G, Enarson, D, Borgdorff, M, Behr, M, & Beyers, N. (2004). Transmission of tuberculosis in a high incidence urban community in South Africa. *International Journal of Epidemiology,* 33(2), 351-357.

Vynnycky, E, Nagelkerke, N, Borgdorff, M, van Soolingen, D, van Embden, J & Fine, P. (2001). The effect of age and study duration on the relationship between

'clustering' of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. *Epidemiology and Infection,* 126(1), 43-62.

Wechkunanukul, K, Grantham, H, Damarell, R, & Clark, R. (2016). The association between ethnicity and delay in seeking medical care for chest pain: a systematic review. *JBI Database System Review Implement Report*, 14(7):208-35.

Weis, S, Pogoda, J, Yang, Z, Cave, D, Wallace, C, Kelley, M, & Barnes, P. (2002) Transmission dynamics of tuberculosis in Tarrant County, Texas. American *Journal of Respriratory Critical Care Medicine*, 166, 36-42.

Wolff, A, & O'Donnell, A. (2004). Pulmonary effects of illicit drug use. *Clinical Chest Medicine*, 25(1), 203-216.

World Health Organization (WHO). (2013). Executive summary. *Global tuberculosis Report,* 2013.1-3.

World Health Organization (WHO). (2011). WHO report 2011: *Global Tuberculosis Control*.

Yeo, I, Tannenbaum, T, Scott, A, Kozak, R, Behr, M, Thibert, L, & Schwartzman, K. (2006). Contact investigation and genotyping to identify tuberculosis transmission to children. *Pediatric Infectious Disease Journal*, 25(11), 1037-1043.

Zolnir-Dovc, M, Poljak, M, Erzen, D, & Sorli, J. (2003). Molecular epidemiology of tuberculosis in Slovenia: results of a one-year (2001) nation-wide study. *Scandinavian Journal of Infectious Disease*, 35(11-12), 8638-8680.

Appendix A: Report of Verified Case of Tuberculosis (RVCT) Form

Patient's Name _____

Street Address _____ (Last) _____ (First) _____ (M.I.) _____ _____ (ZIP CODE)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Centers for Disease Control and Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

FORM APPROVED OMB NO. 0920-0026 Exp. Date 03/31/2017

REPORT OF VERIFIED CASE OF TUBERCULOSIS

1. Date Reported — Month / Day / Year

2. Date Submitted — Month / Day / Year

3. Case Numbers — Year Reported (YYYY) / State Code / Locally Assigned Identification Number

- State Case Number
- City/County Case Number
- Linking State Case Number
- Linking State Case Number

Reason:

4. Reporting Address for Case Counting

City

Within City Limits *(select one)* ☐ Yes ☐ No

County

ZIP CODE — —

5. Count Status *(select one)*

Countable TB Case

☐ Count as a TB case

Noncountable TB Case

☐ Verified Case: Counted by another U.S. area (e.g., county, state)

☐ Verified Case: TB treatment initiated in another country *Specify*_____

☐ Verified Case: Recurrent TB within 12 months after completion of therapy

6. Date Counted — Month / Day / Year

7. Previous Diagnosis of TB Disease *(select one)*

☐ Yes ☐ No

If YES, enter year of previous TB disease diagnosis:

8. Date of Birth — Month / Day / Year

9. Sex at Birth *(select one)*

☐ Male ☐ Female

10. Ethnicity *(select one)*

☐ Hispanic or Latino

☐ Not Hispanic or Latino

11. Race *(select one or more)*

☐ American Indian or Alaska Native

☐ Asian: *Specify*_____

☐ Black or African American

☐ Native Hawaiian or Other Pacific Islander: *Specify*_____

☐ White

12. Country of Birth

"U.S.-born" (or born abroad to a parent who was a U.S. citizen) *(select one)* ☐ Yes ☐ No

Country of birth: *Specify*_____

13. Month-Year Arrived in U.S. — Month / Year

14. Pediatric TB Patients (<15 years old)

Country of Birth for Primary Guardian(s): *Specify*

Guardian 1_____

Guardian 2_____

Patient lived outside U.S. for >2 months? *(select one)* ☐ Yes ☐ No ☐ Unknown

If YES, list countries, *specify*: _____

15. Status at TB Diagnosis *(select one)*

☐ Alive ☐ Dead

If DEAD, enter date of death: Month / Day / Year

If DEAD, was TB a cause of death? *(select one)* ☐ Yes ☐ No ☐ Unknown

16. Site of TB Disease *(select all that apply)*

- ☐ Pulmonary
- ☐ Pleural
- ☐ Lymphatic: Cervical
- ☐ Lymphatic: Intrathoracic
- ☐ Lymphatic: Axillary
- ☐ Lymphatic: Other
- ☐ Lymphatic: Unknown
- ☐ Laryngeal
- ☐ Bone and/or Joint
- ☐ Genitourinary
- ☐ Meningeal
- ☐ Peritoneal
- ☐ Other: Enter anatomic code(s) *(see list)*: 1 2 3
- ☐ Site not stated

Patient's Name _____ State Case No. _____
(Last) (First) (M.I.)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

REPORT OF VERIFIED CASE OF TUBERCULOSIS

17. Sputum Smear *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
Date Collected: Month Day Year

18. Sputum Culture *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
Date Collected: Month Day Year
Date Result Reported: Month Day Year
Reporting Laboratory Type *(select one)*: ☐ Public Health Laboratory ☐ Commercial Laboratory ☐ Other

19. Smear/Pathology/Cytology of Tissue and Other Body Fluids *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
Date Collected: Month Day Year
Enter anatomic code *(see list)*:
Type of exam *(select all that apply)*: ☐ Smear ☐ Pathology/Cytology

20. Culture of Tissue and Other Body Fluids *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
Date Collected: Month Day Year
Enter anatomic code *(see list)*:
Date Result Reported: Month Day Year
Reporting Laboratory Type *(select one)*: ☐ Public Health Laboratory ☐ Commercial Laboratory ☐ Other

21. Nucleic Acid Amplification Test Result *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
☐ Indeterminate
Date Collected: Month Day Year
Date Result Reported: Month Day Year
Enter specimen type: ☐ Sputum
OR
If not Sputum, enter anatomic code *(see list)*:
Reporting Laboratory Type *(select one)*: ☐ Public Health Laboratory ☐ Commercial Laboratory ☐ Other

Initial Chest Radiograph and Other Chest Imaging Study

22A. Initial Chest Radiograph *(select one)*
☐ Normal ☐ Abnormal* (consistent with TB) ☐ Not Done ☐ Unknown
* For ABNORMAL Initial Chest Radiograph: Evidence of a cavity *(select one)*: ☐ Yes ☐ No ☐ Unknown
Evidence of miliary TB *(select one)*: ☐ Yes ☐ No ☐ Unknown

22B. Initial Chest CT Scan or Other Chest Imaging Study *(select one)*
☐ Normal ☐ Abnormal* (consistent with TB) ☐ Not Done ☐ Unknown
* For ABNORMAL Initial Chest CT Scan or Other Chest Imaging Study: Evidence of a cavity *(select one)*: ☐ Yes ☐ No ☐ Unknown
Evidence of miliary TB *(select one)*: ☐ Yes ☐ No ☐ Unknown

23. Tuberculin (Mantoux) Skin Test at Diagnosis *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
Date Tuberculin Skin Test (TST) Placed: Month Day Year
Millimeters (mm) of induration:

24. Interferon Gamma Release Assay for *Mycobacterium tuberculosis* at Diagnosis *(select one)*
☐ Positive ☐ Not Done
☐ Negative ☐ Unknown
☐ Indeterminate
Date Collected: Month Day Year
Test type:
*Specify*_____

25. Primary Reason Evaluated for TB Disease *(select one)*
☐ TB Symptoms
☐ Abnormal Chest Radiograph (consistent with TB)
☐ Contact Investigation
☐ Targeted Testing
☐ Health Care Worker
☐ Employment/Administrative Testing
☐ Immigration Medical Exam
☐ Incidental Lab Result
☐ Unknown

Patient's Name _____ State Case No. _____
(Last) (First) (M.I.)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

REPORT OF VERIFIED CASE OF TUBERCULOSIS

26. HIV Status at Time of Diagnosis *(select one)*

☐ Negative ☐ Indeterminate ☐ Not Offered ☐ Unknown
☐ Positive ☐ Refused ☐ Test Done, Results Unknown

If POSITIVE, enter:
State HIV/AIDS Patient Number: ☐☐☐☐☐☐☐☐☐☐ City/County HIV/AIDS Patient Number: ☐☐☐☐☐☐☐☐☐☐

27. Homeless Within Past Year *(select one)*

☐ No ☐ Yes ☐ Unknown

28. Resident of Correctional Facility at Time of Diagnosis *(select one)* ☐ No ☐ Yes ☐ Unknown

If YES, *(select one)*:
☐ Federal Prison ☐ Local Jail ☐ Other Correctional Facility
☐ State Prison ☐ Juvenile Correction Facility ☐ Unknown

If YES, under custody of Immigration and Customs Enforcement? *(select one)*
☐ No ☐ Yes

29. Resident of Long-Term Care Facility at Time of Diagnosis *(select one)* ☐ No ☐ Yes ☐ Unknown

If YES, *(select one)*:
☐ Nursing Home ☐ Residential Facility ☐ Alcohol or Drug Treatment Facility ☐ Unknown
☐ Hospital-Based Facility ☐ Mental Health Residential Facility ☐ Other Long-Term Care Facility

30. Primary Occupation Within the Past Year *(select one)*

☐ Health Care Worker ☐ Migrant/Seasonal Worker ☐ Retired ☐ Not Seeking Employment (e.g. student, homemaker, disabled person)
☐ Correctional Facility Employee ☐ Other Occupation ☐ Unemployed ☐ Unknown

31. Injecting Drug Use Within Past Year *(select one)*
☐ No ☐ Yes ☐ Unknown

32. Non-Injecting Drug Use Within Past Year *(select one)*
☐ No ☐ Yes ☐ Unknown

33. Excess Alcohol Use Within Past Year *(select one)*
☐ No ☐ Yes ☐ Unknown

34. Additional TB Risk Factors *(select all that apply)*

☐ Contact of MDR-TB Patient (2 years or less) ☐ Incomplete LTBI Therapy ☐ Diabetes Mellitus ☐ Other *Specify* _____
☐ Contact of Infectious TB Patient (2 years or less) ☐ TNF-α Antagonist Therapy ☐ End-Stage Renal Disease ☐ None
☐ Missed Contact (2 years or less) ☐ Post-organ Transplantation ☐ Immunosuppression (not HIV/AIDS)

35. Immigration Status at First Entry to the U.S. *(select one)*

☐ Not Applicable
- "U.S.-born" (or born abroad to a parent who was a U.S. citizen)
- Born in 1 of the U.S. Territories, U.S. Island Areas, or U.S. Outlying Areas

☐ Immigrant Visa ☐ Tourist Visa ☐ Asylee or Parolee
☐ Student Visa ☐ Family/Fiancé Visa ☐ Other Immigration Status
☐ Employment Visa ☐ Refugee ☐ Unknown

36. Date Therapy Started

Month ☐☐ Day ☐☐ Year ☐☐☐☐

37. Initial Drug Regimen *(select one option for each drug)*

	No	Yes	Unk		No	Yes	Unk		No	Yes	Unk
Isoniazid	☐	☐	☐	Ethionamide	☐	☐	☐	Moxifloxacin	☐	☐	☐
Rifampin	☐	☐	☐	Amikacin	☐	☐	☐	Cycloserine	☐	☐	☐
Pyrazinamide	☐	☐	☐	Kanamycin	☐	☐	☐	Para-Amino Salicylic Acid	☐	☐	☐
Ethambutol	☐	☐	☐	Capreomycin	☐	☐	☐	Other	☐	☐	☐
Streptomycin	☐	☐	☐	Ciprofloxacin	☐	☐	☐	*Specify* _____			
Rifabutin	☐	☐	☐	Levofloxacin	☐	☐	☐	Other	☐	☐	☐
Rifapentine	☐	☐	☐	Ofloxacin	☐	☐	☐	*Specify* _____			

Comments:

_____
_____
_____
_____

Patient's Name _____ (Last) (First) (M.I.)

Street Address _____ (Number, Street, City, State) (ZIP CODE)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Centers for Disease Control and Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

FORM APPROVED OMB NO. 0920-0026 Exp. Date 03/31/2017

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Initial Drug Susceptibility Report (Follow Up Report – 1)

Year Counted ☐☐☐☐

State Case Number ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

City/County Case Number ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

Submit this report for all culture-positive cases.

38. Genotyping Accession Number

Isolate submitted for genotyping *(select one)*: ☐ No ☐ Yes

If YES, genotyping accession number for episode: ☐☐☐☐☐☐☐☐☐☐☐☐

39. Initial Drug Susceptibility Testing

Was drug susceptibility testing done? *(select one)* ☐ No ☐ Yes ☐ Unknown

If NO or UNKNOWN, do not complete the rest of Follow Up Report –1

If YES, enter date FIRST specimen collected on which initial drug susceptibility testing was done:

Month ☐☐ Day ☐☐ Year ☐☐☐☐

Enter specimen type: ☐ Sputum

OR

If not Sputum, enter anatomic code *(see list)*: ☐☐

40. Initial Drug Susceptibility Results *(select one option for each drug)*

	Resistant	Susceptible	Not Done	Unknown
Isoniazid	☐	☐	☐	☐
Rifampin	☐	☐	☐	☐
Pyrazinamide	☐	☐	☐	☐
Ethambutol	☐	☐	☐	☐
Streptomycin	☐	☐	☐	☐
Rifabutin	☐	☐	☐	☐
Rifapentine	☐	☐	☐	☐
Ethionamide	☐	☐	☐	☐
Amikacin	☐	☐	☐	☐
Kanamycin	☐	☐	☐	☐

	Resistant	Susceptible	Not Done	Unknown
Capreomycin	☐	☐	☐	☐
Ciprofloxacin	☐	☐	☐	☐
Levofloxacin	☐	☐	☐	☐
Ofloxacin	☐	☐	☐	☐
Moxifloxacin	☐	☐	☐	☐
Other Quinolones	☐	☐	☐	☐
Cycloserine	☐	☐	☐	☐
Para-Amino Salicylic Acid	☐	☐	☐	☐
Other	☐	☐	☐	☐
Specify _____				
Other	☐	☐	☐	☐
Specify _____				

Comments:

_____

_____

_____

_____

_____

Patient's Name _____ (Last) (First) (M.I.)

Street Address _____ (Number, Street, City, State) (ZIP CODE)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Centers for Disease Control and Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

FORM APPROVED OMB NO. 0920-0026 Exp. Date 03/31/2017

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Case Completion Report (Follow Up Report – 2)

Year Counted ☐☐☐☐

State Case Number ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

City/County Case Number ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

Submit this report for all cases in which the patient was alive at diagnosis.

41. Sputum Culture Conversion Documented *(select one)* ☐ No ☐ Yes ☐ Unknown

If YES, enter date specimen collected for FIRST consistently negative sputum culture:

Month ☐☐ Day ☐☐ Year ☐☐☐☐

If NO, enter reason for not documenting sputum culture conversion *(select one)*:

☐ No Follow-up Sputum Despite Induction ☐ Patient Refused ☐ Patient Lost to Follow-Up
☐ No Follow-up Sputum and No Induction ☐ Other *Specify* _____
☐ Died ☐ Unknown

42. Moved

Did the patient move during TB therapy? *(select one)* ☐ No ☐ Yes

If YES, moved to where *(select all that apply)*:

☐ In state, out of jurisdiction *(enter city/county)* *Specify*_____ *Specify*_____
☐ Out of state *(enter state)* *Specify*_____ *Specify*_____
☐ Out of the U.S. *(enter country)* *Specify*_____ *Specify*_____

If moved out of the U.S., transnational referral? *(select one)* ☐ No ☐ Yes

43. Date Therapy Stopped

Month ☐☐ Day ☐☐ Year ☐☐☐☐

44. Reason Therapy Stopped or Never Started *(select one)*

☐ Completed Therapy ☐ Not TB
☐ Lost ☐ Died
☐ Uncooperative or Refused ☐ Other
☐ Adverse Treatment Event ☐ Unknown

If DIED, indicate cause of death *(select one)*:

☐ Related to TB disease ☐ Unrelated to TB disease
☐ Related to TB therapy ☐ Unknown

45. Reason Therapy Extended >12 months *(select all that apply)*

☐ Rifampin Resistance ☐ Non-adherence ☐ Clinically Indicated – other reasons
☐ Adverse Drug Reaction ☐ Failure ☐ Other *Specify* _____

46. Type of Outpatient Health Care Provider *(select all that apply)*

☐ Local/State Health Department (HD) ☐ IHS, Tribal HD, or Tribal Corporation ☐ Inpatient Care Only ☐ Unknown
☐ Private Outpatient ☐ Institutional/Correctional ☐ Other

Comments:

_____
_____
_____
_____

Patient's Name _____ State Case No. _____ REPORT OF VERIFIED CASE OF TUBERCULOSIS
(Last) (First) (M.I.)

Centers for Disease Control and Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

FORM APPROVED OMB NO. 0920-0026 Exp. Date 03/31/2017

REPORT OF VERIFIED CASE OF TUBERCULOSIS

Case Completion Report - *Continued*

(Follow Up Report – 2)

47. Directly Observed Therapy (DOT) *(select one)*

- ☐ No, Totally Self-Administered
- ☐ Yes, Totally Directly Observed
- ☐ Yes, Both Directly Observed and Self-Administered
- ☐ Unknown

Number of weeks of directly observed therapy (DOT) ☐☐☐

48. Final Drug Susceptibility Testing

Was follow-up drug susceptibility testing done? *(select one)* ☐ No ☐ Yes ☐ Unknown

If NO or UNKNOWN, do not complete the rest of Follow Up Report –2

If YES, enter date FINAL specimen collected on which drug susceptibility testing was done:

Month ☐☐ Day ☐☐ Year ☐☐☐☐

Enter specimen type: ☐ Sputum

OR

If not Sputum, enter anatomic code *(see list)*: ☐☐

49. Final Drug Susceptibility Results *(select one option for each drug)*

	Resistant	Susceptible	Not Done	Unknown
Isoniazid	☐	☐	☐	☐
Rifampin	☐	☐	☐	☐
Pyrazinamide	☐	☐	☐	☐
Ethambutol	☐	☐	☐	☐
Streptomycin	☐	☐	☐	☐
Rifabutin	☐	☐	☐	☐
Rifapentine	☐	☐	☐	☐
Ethionamide	☐	☐	☐	☐
Amikacin	☐	☐	☐	☐
Kanamycin	☐	☐	☐	☐

	Resistant	Susceptible	Not Done	Unknown
Capreomycin	☐	☐	☐	☐
Ciprofloxacin	☐	☐	☐	☐
Levofloxacin	☐	☐	☐	☐
Ofloxacin	☐	☐	☐	☐
Moxifloxacin	☐	☐	☐	☐
Other Quinolones	☐	☐	☐	☐
Cycloserine	☐	☐	☐	☐
Para-Amino Salicylic Acid	☐	☐	☐	☐
Other	☐	☐	☐	☐
Specify _____				
Other	☐	☐	☐	☐
Specify _____				

Comments:

_____
_____
_____
_____
_____
_____
_____
_____
_____

Appendix B: Analysis Process Tables

Table B1

Full Stepwise Procedure for Final Model Selection for the Relationship between Tuberculosis Genotype Clustering and Risk Factors

The LOGISTIC Procedure				
Model Information				
Data Set	OUT.TBGENO3			
Response Variable	Clustered			
Number of Response Levels	2			
Model	binary logit			
Optimization Technique	Fisher's scoring			
Number of Observations Read	627			
Number of Observations Used	627			
Response Profile				
Ordered Value	Clustered	Total Frequency		
1	Yes	382		
2	No	245		
Probability modeled is Clustered='Yes'.				
Stepwise Selection Procedure				
Class Level Information				
Class	Value	Design Variables		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1

Substance Abuse	Yes	1
	No	-1
Homelessness	Yes	1
	No	-1
Previous TB Disease	Yes	1
	No	-1
Age	Old	1
	Young	-1
Origin	US-Born	1
	Foreign-Born	-1

Step 0. Intercept entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L = 839.029

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
86.5728	8	<.0001

Step 1. Effect Origin entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	771.455
SC	845.47	780.337
-2 Log L	839.029	767.455

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	71.5741	1	<.0001
Score	72.4848	1	<.0001
Wald	64.4654	1	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
15.902	7	0.026

Note: No effects for the model in Step 1 are removed.

Step 2. Effect Race entered:

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	768.57
SC	845.47	790.775
-2 Log L	839.029	758.57

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	80.4593	4	<.0001
Score	80.6662	4	<.0001
Wald	71.4186	4	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq

6.7964	4	0.147

Note: No effects for the model in Step 2 are removed.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection

Step	Effect Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	Origin		1	1	72.485		<.0001
2	Race		3	2	9.2341		0.026

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Race	3	9.0604	0.0285
Origin	1	11.9979	0.0005

Analysis of Maximum Likelihood Est.s

Parameter		DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.1045	0.114	0.8395	0.36
Race	Other	1	0.3425	0.313	1.1974	0.274
	Hispanic/Latino	1	-0.0617	0.3098	0.0396	0.842
	Black	1	0.1968	0.263	0.56	0.454
Origin	US-Born	1	0.8676	0.2505	11.998	5E-04

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits

Race Other vs. White	2.271	0.764	6.749
Race Hispanic/Latino vs. White	1.516	0.514	4.473
Race Black vs. White	1.963	1.225	3.144
Origin US-Born vs. Foreign-Born	5.67	2.124	15.136

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	49.5	Somers' D	0.354
Percent Discordant	14.1	Gamma	0.557
Percent Tied	36.4	Tau-a	0.169
Pairs	93590	c	0.677

Table B2

All (7) Multivariable Logistic Regression Models including Interaction Terms for the Relationship between Tuberculosis Genotype Clustering and Covariates of Interest

1. The LOGISTIC Procedure Testing the Origin and Race Interaction Term.	
Model Information	
Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1
Origin	US-Born	1		
	Foreign-Born	-1		

Model Convergence Status

Quasi-complete separation of data points detected.

Warning: The maximum likelihood estimate may not exist.

Warning: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	766.743
SC	845.47	802.271
-2 Log L	839.029	750.743

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	88.2858	7	<.001
Score	85.8964	7	<.001
Wald	66.4608	7	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Race	3	2.727	0.4357
Origin	1	0.003	0.9564
Race*Origin	3	1.4567	0.6923

Analysis of Maximum Likelihood Estimates

Parameter		DF	Est.	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-3.30	68.198	0.0023	0.961
Race	Other	1	3.53	68.199	0.0027	0.958

Race	Hispanic/Latino		1	2.50	68.199	0.0013	0.970
Race	Black		1	-2.85	152.5	0.0004	0.985
Origin	US-Born		1	3.72	68.198	0.003	0.956
Race*Origin	Other	US-Born	1	2.86	68.199	0.0018	0.9665
Race*Origin	Hispanic/Latino	US-Born	1	3.62	68.199	0.0028	0.957
Race*Origin	Black	US-Born	1	3.40	152.5	0.0005	0.982

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	49.7	Somers' D	0.359
Percent Discordant	13.8	Gamma	0.565
Percent Tied	36.4	Tau-a	0.171
Pairs	93590	c	0.679

2. The LOGISTIC Procedure Testing the Origin and Previous TB Disease Interaction Term.

Model Information	
Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables		
Origin	US-Born	1		
	Foreign-Born	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1
Previous TB Diagnosis	Yes	1		
	No	-1		

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	768.721
SC	845.47	799.808
-2 Log L	839.029	754.721

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq

Likelihood Ratio	84.3077	6	<.001
Score	83.7743	6	<.001
Wald	73.3824	6	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Origin	1	6.7536	0.009
Race	3	8.5973	0.035
Previous TB Diagnosis	1	2.8822	0.089
Origin*Previous TB Diagnosis	1	0.0238	0.877

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	P > ChiSq
Intercept			1	0.30	0.266	1.3287	0.249
Origin	US-Born		1	0.91	0.3536	6.7536	0.009
Race	Other		1	0.32	0.3146	1.0498	0.305
Race	Hispanic/Latino		1	0.03	0.3111	0.0133	0.908
Race	Black		1	0.19	0.2632	0.5248	0.468
Previous TB Diagnosis	Yes		1	0.45	0.2651	2.8822	0.089
Origin*Previous TB Diagnosis	US-Born	Yes	1	0.04	0.2653	0.0238	0.877

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Race Other vs. White	2.224	0.746	6.632
Race Hispanic/Latino vs. White	1.555	0.526	4.598
Race Black vs. White	1.95	1.215	3.128

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	51.5	Somers' D	0.367
Percent Discordant	14.8	Gamma	0.554
Percent Tied	33.8	Tau-a	0.175
Pairs	93590	c	0.683

3. The LOGISTIC Procedure Testing the Origin and Substance Abuse Interaction Term.

Model Information	
Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables		
Origin	US-Born	1		
	Foreign-Born	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1
Substance Abuse	Yes	1		
	No	-1		

Model Convergence Status

Quasi-complete separation of data points detected.

Warning: The maximum likelihood estimate may not exist.
Warning: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	769.954
SC	845.47	801.041
-2 Log L	839.029	755.954

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	83.0747	6	<.001
Score	82.6179	6	<.001
Wald	71.8622	6	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Origin	1	0.001	0.974
Substance Abuse	1	0.0005	0.981
Race	3	8.3197	0.039
Origin*Substance Abuse	1	0.0006	0.980

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept			1	-2.84	117	0.0006	0.980
Origin	US-Born		1	3.77	117	0.001	0.974
Substance Abuse	Yes		1	2.71	117	0.0005	0.981
Race	Other		1	0.35	0.3129	1.2741	0.259
Race	Hispanic/Latino		1	0.03	0.3102	0.0137	0.9067
Race	Black		1	0.16	0.2638	0.3886	0.533
Origin*Substance Abuse	US-Born	Yes	1	2.92	117	0.0006	0.980

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Race Other vs. White	2.303	0.775	6.845
Race Hispanic/Latino vs. White	1.56	0.528	4.608
Race Black vs. White	1.907	1.188	3.062

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	54.9	Somers' D	0.37
Percent Discordant	17.9	Gamma	0.508
Percent Tied	27.2	Tau-a	0.176
Pairs	93590	c	0.685

4. The LOGISTIC Procedure Testing the Origin and HIV Status Interaction Term.

Model Information	
Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables		
Origin	US-Born	1		
	Foreign-Born	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1
HIV Status	Yes	1		
	No	-1		

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	772.388
SC	845.47	803.475
-2 Log L	839.029	758.388

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq

Likelihood Ratio	80.6409	6	<.000
Score	80.8195	6	<.000
Wald	71.536	6	<.000

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Origin	1	7.6033	0.005
HIV Status	1	0.0589	0.808
Race	3	8.8275	0.031
Origin*HIV Status	1	0.0033	0.954

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.05	0.2409	0.0516	0.820
Origin	US-Born		1	0.87	0.3187	7.6033	0.005
HIV Status	Yes		1	0.05	0.2346	0.0589	0.808
Race	Other		1	0.34	0.3161	1.2119	0.271
Race	Hispanic/Latino		1	-0.06	0.3103	0.043	0.835
Race	Black		1	0.19	0.2638	0.5232	0.469
Origin*HIV Status	US-Born	Yes	1	0.01	0.2339	0.0033	0.954

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Race Other vs. White	2.276	0.761	6.805
Race Hispanic/Latino vs. White	1.507	0.51	4.45
Race Black vs. White	1.945	1.212	3.121

Association of Predicted Probabilities and Observed Responses

Percent Concordant	53.2	Somers' D	0.363
Percent Discordant	17	Gamma	0.517
Percent Tied	29.8	Tau-a	0.173
Pairs	93590	c	0.681

5. The LOGISTIC Procedure Testing the Previous TB Diagnosis and Race Interaction Term.

Model Information

Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables
Origin	US-Born	1

	Foreign-Born	-1		
Previous TB Diagnosis	Yes	1		
	No	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	772.43
SC	845.47	812.398
-2 Log L	839.029	754.43

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.5996	8	<.001
Score	83.8672	8	<.001
Wald	73.1044	8	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Race	3	2.0279	0.566
Origin	1	12.1922	0.001
Previous TB Diagnosis	1	2.1003	0.147
Previous TB Diagnosis*Race	3	0.3187	0.956

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.27	0.2873	0.9274	0.335
Race	Other		1	0.25	0.5204	0.2441	0.621
Race	Hispanic/Latino		1	0.06	0.6315	0.0119	0.913
Race	Black		1	0.34	0.4589	0.5673	0.451
Origin	US-Born		1	0.87	0.2514	12.192	0.000
Previous TB Diagnosis	Yes		1	0.41	0.2873	2.1003	0.147
Previous TB Diagnosis*Race	Yes	Other	1	-0.07	0.4685	0.0239	0.877
Previous TB Diagnosis*Race	Yes	Hispanic/Latino	1	0.11	0.5833	0.0394	0.842
Previous TB Diagnosis*Race	Yes	Black	1	0.16	0.3932	0.1714	0.678

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits		
Origin US-Born vs. Foreign-Born	5.788	2.16	15.50	

Association of Predicted Probabilities and Observed Responses

Percent Concordant	51.5	Somers' D	0.368	
Percent Discordant	14.7	Gamma	0.556	
Percent Tied	33.8	Tau-a	0.176	
Pairs	93590	c	0.684	

6. The LOGISTIC Procedure Testing the Substance Abuse and Race Interaction Term.

Model Information				
Data Set	OUT.TBGENO3			
Response Variable	Clustered			
Number of Response Levels	2			
Model	binary logit			
Optimization Technique	Fisher's scoring			
Number of Observations Read	627			
Number of Observations Used	627			
Response Profile				
Ordered Value	Clustered	Total Frequency		
1	Yes	382		
2	No	245		
Probability modeled is Clustered='Yes'.				
Class Level Information				
Class	Value	Design Variables		
Substance Abuse	Yes	1		
	No	-1		
Origin	US-Born	1		
	Foreign-Born	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0

	Black	0	0	1
	White	-1	-1	-1

Model Convergence Status

Quasi-complete separation of data points detected.

Warning: The maximum likelihood estimate may not exist.
Warning: The LOGISTIC procedure continues in spite of the above warning. Results shown are based on the last maximum likelihood iteration. Validity of the model fit is questionable.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	771.567
SC	845.47	807.095
-2 Log L	839.029	755.567

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	83.462	7	<.001
Score	83.0868	7	<.001
Wald	72.3752	7	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
Substance Abuse	1	1.1265	0.288
Origin	1	11.5907	0.001
Race	3	1.2116	0.750
Substance Abuse*Race	2	0.3685	0.832

Substance Abuse Yes Race Black = Intercept + Substance Abuse Yes + 3 * Race Other - Race Hispanic/Latino - Race Black + 3 * Substance Abuse Yes Race Other - Substance Abuse Yes Race Hispanic/Latino

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.03	0.1871	0.0415	0.838
Substance Abuse	Yes		1	0.16	0.1587	1.1265	0.288
Origin	US-Born		1	0.85	0.2508	11.590	0.000
Race	Other		1	5.39	141.9	0.0014	0.969
Race	Hispanic/Latino		1	5.34	141.9	0.0014	0.97
Race	Black		1	0.18	0.2655	0.474	0.491
Substance Abuse*Race	Yes	Other	1	5.03	141.9	0.0013	0.971
Substance Abuse*Race	Yes	Hispanic/Latino	1	-5.30	141.9	0.0014	0.970
Substance Abuse*Race	Yes	Black	0	0			

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits

Origin US-Born vs. Foreign-Born	5.516	2.064	14.74

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	55	Somers' D	0.372
Percent Discordant	17.8	Gamma	0.512
Percent Tied	27.2	Tau-a	0.177
Pairs	93590	c	0.686

7. The LOGISTIC Procedure Testing the Race and HIV Status Interaction Term.

Model Information	
Data Set	OUT.TBGENO3
Response Variable	Clustered
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring
Number of Observations Read	627
Number of Observations Used	627

Response Profile

Ordered Value	Clustered	Total Frequency
1	Yes	382
2	No	245

Probability modeled is Clustered='Yes'.

Class Level Information

Class	Value	Design Variables		
Origin	US-Born	1		
	Foreign-Born	-1		
Race	Other	1	0	0
	Hispanic/Latino	0	1	0
	Black	0	0	1
	White	-1	-1	-1
HIV Status	Yes	1		
	No	-1		

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	841.029	773.988
SC	845.47	809.515
-2 Log L	839.029	757.988

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	81.0412	7	<.001
Score	81.1609	7	<.001
Wald	71.8016	7	<.001

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq

Origin	1	12.1509	0.000	
Race	3	3.066	0.381	
HIV Status	1	0.4352	0.509	
Race*HIV Status	2	0.3986	0.819	

Race Black HIV Status Yes = Intercept + 3 * Race Other - Race Hispanic/Latino - Race Black + HIV Status Yes + 3 * Race Other HIV Status Yes - Race Hispanic/Latino HIV Status Yes

Analysis of Maximum Likelihood Estimates

Parameter			DF	Est.	SE	Wald Chi-Square	Pr > ChiSq
Intercept			1	0.01	0.2098	0.007	0.933
Origin	US-Born		1	0.87	0.2512	12.150	0.000
Race	Other		1	0.84	0.8407	1.0024	0.316
Race	Hispanic/Latino		1	-0.26	0.528	0.2435	0.621
Race	Black		1	0.16	0.2677	0.3682	0.544
HIV Status	Yes		1	0.12	0.185	0.4352	0.509
Race*HIV Status	Other	Yes	1	0.49	0.7813	0.3984	0.527
Race*HIV Status	Hispanic/Latino	Yes	1	-0.22	0.4951	0.2	0.654
Race*HIV Status	Black	Yes	0	0			

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Origin US-Born vs. Foreign-Born	5.762	2.152	15.42

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	53.2	Somers' D	0.363
Percent Discordant	17	Gamma	0.517
Percent Tied	29.8	Tau-a	0.173
Pairs	93590	c	0.681