Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jennifer Jinju Kim Date

TUBERCULIN SKIN TEST RESULT AND TIME TO SPUTUM CULTURE CONVERSION AMONG PERSONS WITH ACTIVE TUBERCULOSIS

By

Jennifer J. Kim
Master of Public Health

Epidemiology

J. Peter Cegielski

Thesis Field Advisor

Kenneth G. Castro

Faculty Thesis Advisor

TUBERCULIN SKIN TEST RESULT AND TIME TO SPUTUM CULTURE CONVERSION AMONG PERSONS WITH ACTIVE TUBERCULOSIS

By

Jennifer J. Kim

B.S.
Boston University
2014

Faculty Thesis Advisor: Kenneth G. Castro, M.D., F.I.D.S.A.

Thesis Field Advisor: J. Peter Cegielski, M.D., M.P.H.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

ABSTRACT

This retrospective cohort study examines whether tuberculin skin test (TST) results were associated with sputum culture conversion among active tuberculosis (TB) patients in Texas. Previous studies have suggested that TB patients with a negative TST are more likely to develop severe TB disease and have an increased risk of death than those with a positive TST. In this analysis, those with a negative TST were considered exposed, and their culture conversion time, as an indicator of TB treatment success, was compared with that among the positive-TST counterparts. If the patient ever had a positive TST result, defined as 5 mm induration or greater, either before or during his first visit to the study site, he was categorized as a positive TST. Culture conversion was defined as three negative consecutive sputum cultures since the patient's first hospital admission. The study population initially consisted of 961 patients. However, only those with a TST result, at least one baseline positive culture, and at least two cultures overall were eligible for inclusion in the analysis. After exclusions due to unknown TST information and baseline negative or only one culture on record, a total of 491 unique individuals remained. Those who did not convert their sputum culture for any reason were censored at the time of most recent follow-up. The Cox Proportional Hazard model was fit to obtain ratio measures and their confidence intervals. In the crude analysis, negative-TST patients converted slower than positive-TST individuals, but this difference was not statistically significant (HR=0.88; 95% CI: 0.61-1.26). There was a significant multiplicative interaction of malnutrition status and TST on culture conversion time. Among malnourished TB patients, the estimated time of conversion in the negative-TST group was slower than the positive-TST group (aHR=0.58; 95% CI: 0.31-1.07). Among those who were adequately nourished, the negative-TST group converted faster than the positive-TST group (aHR=1.39; 95% CI: 0.87-2.22). These ratio measures were adjusted for age, sex, race, acid-fast bacilli (AFB) smear test, and TB treatment. The differences in time to convert from positive to negative between the two skin test groups were not statistically different even after controlling for covariates.

TUBERCULIN SKIN TEST RESULT AND TIME TO SPUTUM CULTURE CONVERSION AMONG PERSONS WITH ACTIVE TUBERCULOSIS

By

Jennifer J. Kim

B.S.
Boston University
2014

Faculty Thesis Advisor: Kenneth G. Castro, M.D., F.I.D.S.A.

Thesis Field Advisor: J. Peter Cegielski, M.D., M.P.H.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

ACKNOWLEDGEMENT

First and foremost, I would like to thank my brilliant thesis field advisor Peter Cegielski, MD, MPH at Centers for Disease Control and Prevention. He was patient and very generous with his time. Dr. Cegielski allowed this project to be my own work but steered me in the right direction whenever he thought I needed it. I also want to thank my faculty thesis advisor Kenneth Castro, MD, FIDSA at Rollins School of Public Health, Emory University for his guidance. A special thanks to the staff at the University of Texas Health Science Center at Tyler for the data collection including Dr. Cegielski himself, David Griffith, MD, Willis Hassell, Patricia Clark, RN, MSN, Laurie Lackland, RN, Margaret Reed, RN, MSN, Umapraba Pulipaka, MD, David Murphy, Sara Shepherd, MS, and Pattawamadee Chaisrirook.

TABLE OF CONTENTS

CHAPTER I: BACKGROUND

Tuberculosis (TB) is one of the oldest recorded human afflictions. Evidence of TB disease has been detected in Egyptian mummies from 5000 years ago (1). In 1882, German scientist Robert Koch discovered the causative agent of TB, the bacterium called *Mycobacterium tuberculosis*. This airborne infectious disease has been treatable and preventable since the advent of effective antibiotics, starting with streptomycin in 1944, yet continues to be a serious global burden, especially in developing countries where laboratory facilities are limited.

M. tuberculosis is a slow-growing, aerobic bacillus with humans as its primary reservoir (2). TB is spread from person to person, primarily via droplet nuclei released from coughing or sneezing of an individual with pulmonary TB. If another person inhales air containing the infected droplet nuclei, transmission may occur. However, not all persons exposed to an infectious TB patient become infected because transmission depends on how contagious the TB patient was, the environment where exposure occurred, and the length of exposure. Due to its slow doubling time, it takes months to eliminate these mycobacteria once persons with TB disease start effective antibiotic regimens (3).

Although primarily a pulmonary pathogen, *M. tuberculosis* can cause disease throughout the body, not just in the lungs. Following infection, persons can either harbor latent TB infection or progress to develop active TB disease—the distinction between these two states informs the recommended use of specific and effective clinical intervention or treatment. Latent infection is an asymptomatic and non-transmissible state, while patients with active TB disease can exhibit symptoms such as fever, fatigue, weight loss, and lack of appetite. Only those with active pulmonary TB are contagious and commonly have persistent cough and hemoptysis (coughing up of blood) in addition to the general symptoms just described.

There are different diagnostic options for latent TB and active TB disease. As noted earlier, it is crucial to rule out active TB disease before treating latent TB infection because treating active disease with drugs recommended for latent TB could lead to acquired drug resistance, making treatment more

difficult and costly. There are two tests available for detecting latent infection: tuberculin skin test (TST) and interferon-gamma release assays (IGRAs). The TST entails an intradermal injection of commercially available purified protein derivative (PPD). After 48-72 hours, a person with cell-mediated immunity to these antigens will show a delayed-type hypersensitivity reaction, which a clinician can interpret and measure, based on the size of induration and assess the risk of developing active TB disease (e.g. HIV infection, prior history of TB). IGRAs, a more recent testing tool, were introduced with the hope to replace TSTs in the early 2000s (4). They are *in vitro* blood tests of cell-mediated immune response to RD1-encoded antigens that are more specific for *M. tuberculosis* than PPD antigens involved in TSTs. Both the TST and IGRA are widely used but unable to accurately differentiate between latent infection and active disease, a distinction that could be helpful for clinicians to identify patients with increased risk for the progression from infection to disease.

If either TST or IGRA indicates a positive reaction to respective antigens, a medical examination with microbiological tests and radiography are needed to determine whether the person has evidence of TB disease. Diagnosis of TB disease has relied heavily on the presence of *M. tuberculosis* using acid-fast bacilli (AFB) staining of sputum smear microscopy and culture. While AFB smear is quick and easy, it does not confirm a definitive diagnosis of TB—other nontuberculous mycobacteria (NTM) could also appear as positive AFB under microscopic examination (5). Thus, a culture should be prepared for all initial sputum samples to confirm the diagnosis. Cultures, which rely on liquid media such as the Mycobacteria Growth Indicator Tube (MGIT), can require 2 to 5 weeks or even longer for a positive result. Centers for Disease Control and Prevention (CDC) recommends that culture examinations should be completed on all specimens regardless of AFB smear results (6).

In addition to AFB smear and culture tests, nucleic acid amplification test (NAAT) is a more recent diagnostic tool that has been commercially available in the U.S. for the past two decades. Since NAAT offers better accuracy than AFB microscopy and detects *M. tuberculosis* in specimens faster than culture, the CDC has recommended routine use of NAATs as standard practice in the U.S. (7). Other rapid cartridge-based NAATs, such as the GeneXpert MTB/RIF® assay or simply known as GeneXpert,

have also been developed and licensed in recent years. The GeneXpert assay, endorsed by the WHO for use in TB endemic countries, detects DNA of the *M. tuberculosis* complex and genetic mutations associated with resistance to drug rifampin (RIF) in unprocessed sputum (8). Studies have suggested that NAATs could significantly accelerate time to diagnosis and even time-to-treatment initiation among TB patients. In Peralta et al.'s study, the median time-to-treatment initiation from NAAT report data was only one day for NAAT-positive patients, much shorter than 11 days for NAAT-negative patients (7). Similar to this time difference between NAAT-negative and NAAT-positive patients, our investigation compares time to sputum culture conversion between TST-negative and TST-positive individuals. Persons with active TB and a negative tuberculin skin test result traditionally have delayed diagnosis of TB and treatment initiation because their false-negative result could have misled clinicians to think that the patient did not have TB based on the skin test alone.

Standard treatment for drug-susceptible TB disease consists of four first-line antimicrobials including isoniazid, rifampin, pyrazinamide and ethambutol. They are relatively cheap, safe and highly effective—this recommended standardized treatment regimen can cure more than 95% of patients with drug-susceptible TB (9). However, resistance to some of the first-line drugs (e.g. isoniazid) is relatively common and can be tested using culture-based methods. Multidrug-resistant TB (MDR-TB) is defined as *M. tuberculosis* strains resistant to at least isoniazid and rifampin and can still be treated with second-line drugs with reasonably high cure rates. However, these second-line drugs are more toxic and costlier than first-line drugs. In the early 1990s, clinicians had to use second-line drugs more frequently due to the emergence of MDR-TB. This growing resistance to second-line drugs later gave rise to extensively drug-resistant TB (XDR-TB) (10). XDR-TB is not only resistant to isoniazid and rifampin but also to fluoroquinolones and any of the three injectable second-line aminoglycosides. Both over- and improper use of antibiotics further complicates drug resistant TB, requiring complex, toxic and costly regimens (11). Therefore, early diagnosis and immediate initiation of an effective treatment against drug-resistant TB are crucial for optimizing treatment outcomes and preventing disease transmission.

The World Health Organization (WHO) estimates that 10.4 million incident cases of TB and 1.4 million TB deaths occurred worldwide in 2015 (12). In general, declining trends in TB incidence, prevalence and mortality have been observed over the last decade. However, global elimination of the disease is still beyond reach, and efforts to reduce TB burden clash with economic challenges. The epidemiology of TB reveals stark disparities between developed and developing countries. For instance, TB incidence is about 300-fold higher in South Africa than in the United States (4). South Africa also has the highest burdens of drug-resistant TB in the world—the number of XDR-TB has increased by a factor of 10 in the past decade, to about 1500 cases in 2012 (11). In western Europe and the U.S., on the other hand, TB-related deaths fell by greater than 100-fold, attributed to improvements in nutrition and socioeconomic conditions (4). TB is a poverty-related disease that disproportionately affects the poorest, most vulnerable and marginalized populations, and improving access to diagnosis and care for these persons has been particularly challenging.

People are at high risk for TB infection if they have close contact with a known or suspected TB case, were born in areas where TB is prevalent, reside or work in high-risk congregate settings (e.g. prison, hospital), or use illicit injectable drugs (11). Among risk factors for progression from latent infection to active TB disease, human immunodeficiency virus (HIV) infection is the strongest. People with HIV/AIDS are at extremely high risk of TB because they are immunocompromised. Globally, 12% of all new active TB disease cases and 25% of all TB-related deaths have been observed in HIV-positive individuals, and the majority of these HIV-associated TB disease cases and TB-related deaths occurs in Africa (4). Furthermore, TB/HIV co-infected patients struggle with treatment—73% success rate compared to 87% in HIV-uninfected patients (13). Most public health organizations recommend that all TB/HIV co-infected patients receive antiretroviral therapy (ART) immediately, but the coverage of ART among HIV-infected persons remains relatively low in many African countries. Thus, expanding the coverage of ART in HIV-infected TB patients is the utmost priority in areas with a high burden of both diseases.

A national surveillance system for TB cases in the U.S. was established by the CDC in 1953 (14). The TB case definition for surveillance relies on both laboratory and clinical criteria—a case is confirmed by identification of *M. tuberculosis* from a clinical specimen or by demonstration of AFB when a culture could not be obtained. With the introduction of antimicrobial therapy in 1950, the U.S. saw the downward trend in TB morbidity between 1953 and 1985: the incidence rate declined from 53 to 9.3 cases per 100,000 persons (14). However, the trend reversed in the late 1980s, and the incidence rate reached a peak of 10.5 cases per 100,000 persons in 1992. This sudden increase was correlated with onset of the acquired immunodeficiency syndrome (AIDS) epidemic; the geographic and demographic distributions of the two epidemics were similar during the same period (15). When these concerns were recognized in the early 1990s, the government increased federal funding and implemented stronger TB control efforts. Since then, the incidence of TB cases decreased annually and reached a modest rate of 5.8 per 100,000 persons in 2000 (14). According to the most recent surveillance report by CDC, the TB incidence rate in the U.S. has remained relatively stable at about 3.0 per 100,000 persons since 2013 (16).

In addition to HIV/AIDS, the rise of TB cases in the U.S. between 1985 and 1992 was attributed to the development of multidrug-resistant TB and increased numbers of foreign-born individuals who moved from TB-endemic countries (15). The influx of immigrants in particular posed a serious challenge to reducing TB morbidity in the U.S. The countries that accounted for the majority of TB cases among foreign-born persons in the U.S. include Mexico, Philippines, Vietnam, India, China, Haiti, and Korea (14). Also, the epidemiology of TB in foreign-born persons varies by state. In Texas, the focus of this investigation, 626 (31%) of 1,992 TB cases occurred among foreign-born individuals in 1997, which was slightly below 39% foreign-born cases in the U.S. overall (19,851 total) during that year (17). Among these foreign-born TB cases in Texas, more than half (358 out of 626 total) reported Mexico as their country of origin. Today TB continues to disproportionately affect foreign-born individuals, accounting for more than 60% of total cases in the U.S. (16).

CHAPTER II: MANUSCRIPT

Tuberculin Skin Test Result and Time to Sputum Culture Conversion among Persons with Active Tuberculosis

Jennifer J. Kim, J. Peter Cegielski

Abstract

Background: Tuberculin skin test (TST) has been widely used to detect latent tuberculosis (TB) infection, yet some people with active TB disease do not respond to tuberculin and test false-negative. Previous studies have suggested that TB patients with certain risk factors or immunocompromising conditions are more likely to have a negative TST and an increased risk of death or worse disease manifestations compared to relatively healthier counterparts.

Methods: To extend these findings, we analyzed a database consisting of all TB patients who were hospitalized in the University of Texas Health Science Center at Tyler from 1985 to 2010. Patients without TST and eligible culture results were excluded from analysis. The Cox proportional hazard method was used to compare the number of days to initial sputum culture conversion between people with negative and positive TST results.

Results: Of 961 TB patients, 491 persons were eligible for inclusion in the analysis. Persons with a negative TST accounted for 15% of the culture converters. There was a significant multiplicative interaction between malnutrition and TST with respect to culture conversion. Among malnourished TB patients, the negative-TST group converted slower than the positive-TST group (adjusted hazard ratio 0.58, 95% confidence interval 0.31-1.07). Within adequately nourished individuals, the negative-TST group converted sooner than the positive-TST group (adjusted hazard ratio 1.39, 95% confidence interval 0.87-2.22).

Conclusion: A negative TST was not significantly associated with time to culture conversion in this hospitalized TB cohort, even after adjustment for age, sex, race, acid-fast bacilli (AFB) status, and TB treatment.

Introduction

The tuberculin skin test (TST) is a common diagnostic tool for infection with *Mycobacterium tuberculosis* complex. This intradermal injection method has been widely used to identify individuals with latent tuberculosis (TB) so that they could receive appropriate treatment and prevent progression to TB disease. Normally, a positive TST result, based on measurable skin induration, indicates a cell-mediated immune response to mycobacterial antigens called purified protein derivative (PPD). However, some active TB disease patients do not respond to PPD and falsely test negative on the TST. This absence of delayed-type hypersensitivity (DTH) responses of host immune system memory T cells to antigens is defined as "anergy." It is estimated that up to 25% of persons with active TB may be anergic, or have no reaction to PPD tuberculin (2). Thus, a negative TST result does not necessarily rule out TB disease in a person with symptoms of the disease, and anyone who is suspected to have active TB should have a full diagnostic evaluation.

It is well recognized that the host immune response is an important determinant of the clinical presentation of active TB disease and the extent to which a TB patient is going to react to PPD. Epidemiological studies by Auld et al. (18, 19) found a significant association between TST result and death or disseminated disease. The authors highlight that people with immunocompromising conditions including human immunodeficiency virus (HIV) infection and malnutrition are more likely to have a negative TST and even a greater risk of death or disseminated disease compared to immunocompetent individuals. Such conditions interfere with cell-mediated DTH responses to PPD. Differences in the cytokine profiles of TST-negative and TST-positive TB patients have also been observed in biomedical studies. Anergic, TST-negative individuals demonstrated antigen-specific impaired T helper cells that no longer effectively produce cytokines such as interleukin (IL)-2, IL-10, and interferon (IFN)-γ (20, 21). These observations suggest that patients with a negative TST may lack in adequate immune response to the mycobacterium, and they may continue to struggle with TB disease even with appropriate treatment (18). Therefore, in addition to diagnostic utility, TST may have prognostic value among TB patients.

A definitive diagnosis of TB requires confirming the presence of the mycobacterium in a patient's bodily secretions or tissues. Sputum cultures have been the gold standard for laboratory confirmation of pulmonary TB disease, but beyond their diagnostic use, cultures serve as an early endpoint in phase II clinical trials of TB treatment (22). Among individuals with pulmonary TB, conversion of positive sputum cultures to negative, or no growth of *M. tuberculosis*, is an important indicator of treatment response and declining infectiousness. For patients with multidrug-resistant TB (MDR-TB) in particular, clinicians often determine the overall duration of MDR-TB treatment on the basis of culture conversion (23).

For patients with pulmonary TB and positive pretreatment sputum, the response to TB treatment should be evaluated by repeated examinations of sputum specimens at monthly intervals until conversion of sputum is documented by three consecutive negative sputum specimens (24). Sputum conversion should occur among these patients by two months of treatment with regimens containing isoniazid and rifampin. If conversion does not occur after 5 to 6 months, treatment is considered to fail. In case of treatment failure, susceptibility tests should be performed so that the treatment regimen could be adjusted accordingly. An individual who successfully completes the treatment regimen, with negative cultures at the end of treatment, is considered cured of TB. Some risk factors have shown to hinder patients from achieving sputum culture conversion: baseline positive smear, alcoholism, resistance to certain TB drugs, and poor outcome of previous anti-TB treatment (25). In general, effective treatment is crucial to TB control and prevention, and the monitoring of the timeliness of TB treatment is necessary to achieve these goals.

Even though TST reactivity has been widely studied in the context of latent TB infection, mortality, and clinical presentation of TB disease, we have not come across any large studies that specifically looked at the difference in time to sputum culture conversion by TST result. To further characterize the relationship between TST and TB treatment outcomes, we compared the time to conversion from positive to negative sputum culture growth among TST-negative and TST-positive patients using a clinical TB database from 1985-2010.

Methods

Specific Aim

The aim of this study was to compare time to initial sputum culture conversion between negative and positive TST patients in order to determine if negative TST is a significant predictor of slower sputum culture conversion from positive to negative after their first admission to the study site.

Hypothesis

The null hypothesis was that there would be no statistical difference in time to initial sputum culture conversion between negative-TST and positive-TST patients who were hospitalized with active TB. The alternative hypothesis was that there would be a meaningful difference in time between the two TST groups.

Study design

This was a retrospective cohort study of hospital-based TB patients in Texas, United States.

Study population

The sampling frame consisted of all patients who were admitted to the University of Texas Health Science Center (UTHSC) at Tyler between January 1, 1985 and December 31, 2010 with a diagnosis of TB (ICD-9-CM codes 010-018 or 137). A 25% simple random sample was selected, except for 100% sampling of three "complicated" groups: drug-resistant TB, HIV-positive, and those whose serum drug levels were measured. As a result, 961 patients were included in the original dataset before we excluded certain individuals who did not meet the inclusion criteria for this specific analysis.

Administrative, sociodemographic, medical history, and present admission clinical data for the selected patients were obtained from their medical records. Since patients with extrapulmonary TB often do not have follow-up cultures, the cohort was restricted to those with pulmonary TB. Not all patients were able to achieve sputum culture conversion or had TST results on record. Therefore, a total of 470 patients were omitted from further analysis because 231 of them had baseline negative cultures or only one culture, and other 239 individuals did not have TST (Figure 1). The final analysis dataset consisted of 491 eligible TB patients.

The parent study was approved by ethics committees at UTHSC at Tyler and Centers for Disease Control and Prevention, and the present secondary analysis of these data was exempt from further Institutional Review Board approval by Emory University's ethics committee.

Definitions

The main predictor was the patient's TST result from either before or at first admission to the study site. A transverse diameter of induration less than 5 mm would indicate that the person did not react to PPD in the skin test and was hence considered as TST-negative. Otherwise, the person was TST-positive if the diameter was greater than or equal to 5 mm. A collective TST variable was created to summarize all tuberculin skin tests performed on each individual. If the patient ever had a positive TST result from before or during the first hospital admission, the person was considered TST-positive; otherwise, the person was classified as TST-negative.

The outcome of interest was time from admission to negative sputum culture conversion. Mycobacteriology cultures were performed for each subject at least once a week throughout his or her stay in the hospital. Each test was coded as positive (any growth of *M. tuberculosis*), negative (no mycobacterial growth), or other (contaminated or non-TB mycobacteria). For the purpose of this study, only patients with more than one culture were analyzed because there could be no ascertainment of conversion in patients with only one culture. Those who had only two or three cultures in total were considered to have conversion if the first culture was positive, and the last culture was negative (Figure 1). Those who did not convert their sputum culture for any reason were censored. The time to censoring was the interval (in number of days) between the date when TB treatment started and the date of discharge or last contact with the patient on record.

Covariates

The following sociodemographic factors were assessed: age, sex, race/ethnicity, birthplace, usage of intravenous and illegal drugs, alcohol consumption, tobacco smoking, homelessness, and health insurance coverage. Clinical factors included HIV status, nutritional status (indicated by below or over the body mass index [BMI] of 18.5), acid-fast bacilli (AFB) smear test, multidrug resistance, and TB

treatment consisted of either first line drugs or second line drugs (taken for at least 30 days since the patient's first admission to the hospital).

Statistical analysis

The cohort was weighted to reflect the sampling probabilities as described above and stratified based on the inclusion criteria. The crude association between TST and the dichotomous culture conversion variable (converted vs. not converted) was evaluated using the Pearson chi-square test. Each covariate was screened for its association with TST or culture conversion. If not associated at α=0.05 with both the exposure and conversion based on the chi-square test, that variable was not considered for further analysis. Some predictors were associated with either TST or culture conversion but not both. In that case, they were assessed by *a priori* criteria or comparing the adjusted and crude estimates from logistic regression models (SAS version 9.3 Proc Surveylogistic). If there was more than 10% change in estimates after adjusting for it, the variable was included in the initial Cox proportional hazards model in case of potential confounding of that certain factor.

The time-to culture conversion in negative and positive TST groups were plotted using Kaplan-Meier techniques. For the comparison of rates between the two groups, the log-rank test was performed. The proportional hazards (PH) assumptions were checked using graphical techniques (the log-log survival curves), extended Cox models with time-dependent covariates, and Schoenfeld's goodness-of-fit test. Almost all of the predictors except foreign birth did not grossly violate the PH assumptions, so a regular Cox PH model was fit without accounting for foreign birth. Also, the hazard ratios for HIV status and multidrug-resistant TB were not obtained because these variables were stratification criteria for 100% sampling and hence were integrated in the Stratum statement instead of Model statement in SAS Proc Surveyphreg.

Interaction terms between covariates and TST were first examined using an overall likelihood ratio test, comparing the full model that contained all interaction terms and the reduced model without them. If the null hypothesis of no difference between -2 log likelihoods for these two models was rejected, the interaction terms were examined by backwards elimination using the Wald chi-square

statistic. The model was refit each time the most insignificant interaction term was dropped. This process was repeated, dropping one interaction term at a time, until only significant (at α=0.05) terms remained in the model. Covariates were then assessed for confounding using the all-possible subsets approach. The hazard ratios for all strata of the interaction term of the models with and without adjusting for a particular covariate were compared. If the hazard ratio for the main exposure was meaningfully different or outside the 10% range of the gold standard hazard ratio, the covariate was retained in the model as a confounder. The final adjusted proportional hazard ratio was obtained from a model that contained conversion time as the dependent variable, TST as the main predictor, and important covariates or potential confounders. Once the final Cox Proportional Hazard model was determined, adjusted survival curves were plotted.

All data were analyzed by using SAS for Windows, version 9.3 (SAS Institute, Inc., Cary, North Carolina).

Results

The study cohort consisted of 961 TB patients who were hospitalized at UTHSC at Tyler from 1985 to 2010, but a total of 491 met the inclusion criteria and remained in the final analysis dataset. Among these 491 patients, 88 (18%) had a negative TST result, and 403 (82%) had a positive TST (Table 1). There were 278 (57%) sputum culture converters, while the other 213 (43%) patients did not convert at all or were censored. The study population was weighted to reflect the sampling probabilities. These weighted data represented 1,333 individuals: 228 (17%) were TST-negative, 1,105 (83%) were TST-positive, 700 (53%) converted, and 633 (48%) did not convert or were censored.

There were more men (75%) than women (25%), and about 70% of the patients were neither too young nor old, falling in the 25 to 44 years of age group. The racial/ethnic distribution of this cohort was more commonly White/Caucasian (38%), followed by Blacks (35%), Hispanics (22%), and Asians (3%) as the smallest racial makeup. One third of the study population (33%) were born outside the U.S., which reflected the percentage of foreign-born TB cases in Texas at the time, as reported by CDC in 1997 (12). Over half (52%) indicated that they were current smokers at the time of their first hospital admission, but

the patients were more modest with self-reported alcohol consumption—45% of them said they never drink. Only a small fraction (6%) of patients said they used intravenous drugs, but a greater proportion (12%) admitted to illegal drug use. There were some notable clinical risk factors that the patients were already struggling with before they came into the hospital: 28% were malnourished (i.e. BMI less than 18.5), 79% were AFB smear-positive, 41% had cancer, 12% had diabetes mellitus, and 17% had a prior history of TB. Almost all patients received appropriate TB treatment for at least 30 days since their first admission to the study site: 1,257 (94%) and 468 (35%) received first line and second line drugs, respectively. Relatively few patients, less than 5% of the cohort, were HIV-positive or had multidrug-resistant TB.

Certain sociodemographic and clinical characteristics were compared by TST result and by culture conversion. According to the Pearson Chi-square test, age ($p=0.02$), HIV status ($p=0.002$), malnutrition ($p=0.01$), AFB smear test ($p=0.001$), and second line drugs (0.02) were significantly associated with TST (Table 2). In each level or category of all risk factors analyzed, there were more (greater than 50%) TST-positive than TST-negative individuals. In the bivariate analysis of the weighted study population using a dichotomous outcome of sputum culture conversion, TST was not significantly associated with conversion ($p=0.31$) (Table 3). There were 107 (weighted count) converters with a negative TST result, which accounted for 15% of 700 converters in the weighted sample overall; conversely, the converters accounted for 47% of TST-negative patients. Most of the selected risk factors except HIV status and birthplace were significantly associated with culture conversion: age ($p=0.0001$), sex ($p<0.0001$), race ($p=0.02$), illegal drug use ($p=0.001$), alcohol consumption ($p<0.0001$), tobacco smoking ($p<0.0001$), homelessness ($p<0.0001$), malnutrition ($p=0.0002$), AFB smear test ($p=0.02$), multidrug resistant-TB ($p=0.002$), and TB treatment including first line drugs ($p=0.03$) and second line drugs ($p=0.009$). Birthplace was neither associated with TST nor sputum culture conversion in this analysis, while AFB smear status was associated with both the exposure and the binary outcome.

When the crude, bivariate Cox Proportional Hazard model was used, not adjusting for any other predictors, negative-TST patients were less likely to convert their sputum culture compared to positive-

TST patients (Hazard Ratio (HR)=0.88, 95% Confidence Interval (CI): 0.61-1.26) (Table 4). The Kaplan-Meier curves for the negative and positive-TST groups reflected this crude association: the log-rank test did not indicate that the curves differed significantly by TST result (p=0.61) (Figure 2). According to the unadjusted model of each risk factor, the youngest (HR=0.72, 95% CI: 0.40-1.28), female (HR=0.70, 95% CI: 0.46-1.02) and AFB smear-positive (HR=0.50, 95% CI: 0.29-0.85) patients were less likely to convert their culture than their respective referent groups. Among these risk factors, AFB smear test was the only significant predictor of culture conversion.

The likelihood ratio test suggested a significant difference between the full model with interaction terms and the reduced model without them. After dropping insignificant terms, only the interaction between TST and malnutrition remained significant (Wald chi-square statistic=29.3; p=0.0001) (tables not shown). Due to this significant multiplicative interaction between TST and malnutrition with respect to culture conversion, two separate ratio measures and their confidence intervals were obtained for the malnourished and the adequately nourished. There were no meaningful changes (within 10%) when potential confounders were dropped from the gold standard model. Hence, the final Cox proportional model included TST the main predictor, age, sex, race, AFB smear status, and TB treatment (first and second line drugs) as covariates.

When adjusted for age, sex, race, AFB smear status, and treatment (first line drugs or second line drugs), malnourished TB patients with negative TST converted slower than their positive-TST counterparts (aHR=0.58; 95% CI: 0.31-1.07) (Table 5). However, adequately nourished patients showed just the opposite: as long as the patients had sufficient nourishment, the negative-TST group converted sooner than the positive-TST group (aHR=1.39; 95% CI: 0.87-2.22). The survival curves adjusted for mean values of age, sex, race, acid-fast bacilli smear status, and TB treatment illustrate these ratio findings. In the adjusted survival curves for malnourished TB patients, the negative-TST group, denoted by a red curve, is above the positive-TST curve in blue (Figure 3). In the adequately nourished plot, on the other hand, the positive-TST curve is above the negative-TST curve (Figure 4).

Discussion

Both bivariate and multivariate analyses did not suggest any statistically significant association between tuberculin skin test results and sputum culture conversion. Therefore, we failed to reject the null hypothesis that there would be no statistical difference in time to initial sputum culture conversion between negative-TST and positive-TST patients.

Contrary to what previous studies have alluded (18, 19), birthplace was not associated with TST in this analysis. However, no significant relationship between birthplace and sputum culture conversion has been observed in both our own and another U.S. study by Liu et al. from 1999 (26). Currently, there are limited data regarding the difference in time to convert between U.S.-born and foreign-born TB patients. While birthplace was neither associated with TST nor sputum culture conversion, AFB smear test was significantly associated with both TST and culture conversion in our study, consistent with Auld et al. (19) and Kurbatova et al.'s findings (25).

After controlling for age, sex, race, AFB smear status, and treatment with first line drugs and second line drugs, we observed that malnourished individuals who were hospitalized in UTHSC at Tyler with active TB and a negative TST result were slower to convert their cultures than malnourished patients with a positive TST result. In contrast, among those who were adequately nourished, in contrast, patients with a negative TST result converted to negative cultures faster than positive-TST counterparts. However, such differences in culture conversion time between the two TST groups were not statistically significant regardless of nutritional status.

Malnutrition is an important risk factor for TB because cell-mediated immunity (CMI), the main host defense against TB, is diminished by malnutrition. Protein-energy undernutrition, for instance, impairs T-cell-mediated immunological defenses, increasing the risk of infectious diseases in general and TB in particular (27, 28). Malnourished individuals are also at higher risk of progressing from latent TB infection to active disease due to the close link between malnutrition and a deficit of CMI (29). The geographic overlap between TB incidence and prevalence of undernutrition reflects this interaction

between malnutrition and CMI in developing countries: populations at highest risk for poor nutrition are also at high risk for TB, poverty being the common denominator (29).

Our findings were partially consistent with other studies where malnutrition hindered TB patients from recovering or achieving treatment success. Kurbatova et al. (22, 25) observed that malnourished individuals were less likely to have culture conversion (odds ratio= 0.37, 95% CI: 0.28-0.50; HR=0.87, 95% CI: 0.72-0.93). In this study, the association between TST results and culture conversion differed between malnourished and adequately nourished individuals. Among malnourished or underweight patients, who made up 28% of the study cohort, negative TST results were associated with slower culture conversion. This result fits a conceptual model in which malnutrition weakens CMI, reflected by a negative TST, and consequently, weaker CMI leads to less vigorous killing of mycobacteria despite appropriate TB treatment. In contrast, among adequately nourished patients, TST-negative individuals converted faster than those with a positive TST, even though this difference in time to convert was not statistically significant. In the latter case, the effect of negative TST, which supposedly represents weak CMI or disease severity, may be partially independent from the effects of nutrition. Or rather, the effects of negative TST or nutrition may be additive, operating in parallel instead of being linked through CMI.

Although sputum culture conversion may serve as a fair indicator of how well a person responds to TB treatment, there is an important distinction between prevention of TB among healthy people who do not have TB yet and response to TB treatment among active TB patients. Based on other relevant data available, nutrition seems to be more correlated with prevention than it does with response to treatment. Compared to persons with normal BMI, undernourished individuals have an increased risk of developing TB, while obese individuals have a lower risk of TB (27). In theory, intensive nutritional support would help patients recover from active TB, but evidence from controlled intervention studies do not support this (30). While it is true that TB patients with very low BMI and serum albumin levels are more likely to have worse TB disease and outcomes, further research is warranted to examine how nutritional supplementation may improve the patient's responsiveness to TB treatment, as indicated by how fast he or she achieves culture conversion, if such intervention works successfully.

Limitations

An extensive collection of sociodemographic and clinical characteristics of TB patients was available in hospital records. With this information, the time to convert from a positive to negative sputum culture could be compared between negative-TST and positive-TST groups while controlling for other risk factors of TB, such as AFB smear test and TB treatment, that could potentially influence the TST or culture conversion. Despite such strengths, the study carries some noteworthy limitations.

First of all, the study only includes hospitalized patients who were more likely to have severe disease than TB patients in general. This cohort experienced some dramatic demographic shifts in TB populations in the U.S. between 1985 and 2010. Up to the mid-1990s, virtually all TB patients were admitted to the hospital, but by 2010, only those who were very ill were hospitalized. Therefore, the difference in time to sputum culture conversion among the hospitalized cohort by tuberculin skin test result is not generalizable to the non-hospitalized population of TB patients.

Due to exclusions based on the inclusion criteria described above, the study sample was restricted to only 491, about half of the original cohort of 961 hospitalized TB patients. Since these criteria were based on the availability of TST results and the patient's ability to convert, the study population is not quite representative of patients who might have had the skin test or a history of TB prior to their first admission to UTHSC at Tyler. During their diagnostic evaluation process in the hospital, clinicians might have skipped administering the TST if they knew that the patient already had the tuberculin skin test before they arrived at the study site. Patients with only one culture test performed during their entire hospital stay were also excluded—these individuals might have recovered faster and left the hospital earlier than those who stayed longer and hence were more likely to perform more laboratory tests.

More comprehensive and prospective measurement of the main exposure should be performed in future studies. In this analysis, the patient was considered TST-positive if he or she ever had a positive TST result on record. More specifically, there were three different PPD skin test variables in the dataset that we used to summarize a patient's overall TST result: two (one being continuous skin induration measurement and the other coded as binary) obtained from the patient's first admission to the UTHSC at

Tyler and one dichotomous TST variable taken from a previous health care provider. Any positive result out of these three skin tests prevailed, making the patient TST-positive regardless of preceding or subsequent negative TST results. The patient could have tested positive and later negative, or vice versa. The timing of the skin test should be taken into account for this type of analysis because in some persons who are tested many years after initial TB infection, they may show a negative reaction (2). However, the skin test can boost their ability to react to PPD and later produce a positive reaction to subsequent tests. Data on the specific timing of these various TSTs relative to starting treatment were not yet available at the time of this analysis. Also, TST could be further categorized into multiple levels instead of simply lumping three PPD tests into binary (negative or positive) results. With different levels of TST, the effect of increasing TST induration size on conversion time could be analyzed by fitting a multinomial logistic regression.

Among TB patients with a negative TST result, the negative result might have influenced clinicians, who are prone to relying on the TST as an adjunctive diagnostic test for TB disease. Thus, the negative TST could have misled such health providers into thinking that the person did not have TB, at least initially. For these false-negative individuals, treatment could have been delayed, further complicating their disease. The effect of treatment delay on culture conversion time could have been in the same direction as an immunological effect—a weaker immune response to TB leading to slower culture conversion. In other words, the delay in treatment could have confounded the observed association between TST and culture conversion. It is possible that this underlying effect could enhance the findings of Auld et al. (18, 19) that negative TST was associated with worse disease manifestations and higher mortality. For a more convoluted analysis, time between the skin test and diagnosis of TB or initiation of TB treatment should be calculated and controlled for as a covariate in the Cox Proportional Hazard model.

Ideally, an extended Cox model should be fit to account for any time-dependent covariates. In this analysis, foreign birth failed all three tests for checking the Proportional Hazards (PH) assumption. Due to limited abilities of “Proc Surveyphreg” in SAS, we decided to omit foreign birth from the model

and obtained ratio measures that did not account for the possible confounding effect of foreign birth on the relationship between tuberculin skin test result and sputum culture conversion. Usually, covariates that do not satisfy the PH assumption are added to the "Stratum" statement within "Proc Phreg" to produce different baseline hazard functions in multiple strata of each time-dependent covariate. The Stratum statement of Proc Surveyphreg rather concerns the sample survey design. Unfortunately, there was no equivalent function in Proc Surveyphreg that could account for both the study's stratified sample design and time-dependent covariates such as foreign birth.

REFERENCES

1. Daniel TM. The history of tuberculosis. Respiratory Medicine, 2006; 100: 1862-1870.
2. Heymann DL. Control of communicable diseases manual. American Public Health Association, 2004.
3. Centers for Disease Control and Prevention. Extensively drug-resistant TB. 2016; (https://www2c.cdc.gov/podcasts/player.asp?f=8644444). (Accessed January 15, 2017).
4. Pai M, Behr MA, Dowdy D, Dheda K, Divangahi M, Boehme CC, Ginsberg A, Swaminathan S, Spigelman M, Cetahun H, Menzies D, Raviglione M. Primer: Tuberculosis. Nature, 2016.
5. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, Holland SM, Horsburgh R, Huitt G, Iademarco MF, Iseman M, Olivier K, Ruoss S, von Reyn CF, Wallace RJ Jr, Winthrop K. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. American Journal of Respiratory and Critical Care Medicine, 2007; 175(4): 367-416
6. Centers for Disease Control and Prevention. TB elimination: diagnosis of tuberculosis. Atlanta: Centers for Disease Control and Prevention, 2011.
7. Peralta G, Pennan B, Pascopella L. Use of nucleic acid amplification tests in tuberculosis patients in California, 2010-2013. Open Forum Infectious Diseases, 2016; 3(4): ofw230.
8. Centers for Disease Control and Prevention. Availability of an assay for detecting Mycobacterium tuberculosis, including rifampin-resistant strains, and considerations for its use—United States, 2013. Morbidity and Mortality Weekly Report (MMWR), 2013; 62(41).
9. Centers for Disease Control and Prevention. Emergence of extensively drug resistant tuberculosis. 2007; (https://www2c.cdc.gov/podcasts/player.asp?f= 4733). (Accessed January 15, 2017).

10. Shah NS, Wright A, Bai GH, Barrera L, Boulahbal F, Martin-Casabona N, Drobniewski F, Gilpin C, Havelkova M, Lepe R, Lumb R, Metchock B, Portaels F, Rodrigues MF, Rusch-Gerdes S, Van Deun A, Vincent V, Laserson K, Wells C, Cegielski JP. Worldwide emergence of extensively drug-resistant tuberculosis. Emerging Infectious Diseases. 13(3): 380-387.
11. Shah NS, Auld SC, Brust JCM, Mathema B, Ismail N, Moodley P, Mlisana K, Allana S, Campbell A, Mthiyane T, Morris N, Mpangase P, van der Meulen H, Omar SV, Brown TS, Narechania A, Shaskina E, Kapwata T, Kreiswirth B, Gandhi NR. Transmission of extensively drug-resistant tuberculosis in South Africa. The New England Journal of Medicine, 2017; 376: 243-253.
12. World Health Organization. Global tuberculosis report 2016. Geneva: WHO; 2016.
13. Sulis G, Roggi A, Matteelli A, Raviglione MC. Tuberculosis: epidemiology and control. Mediterranean Journal of Hematology and Infectious Diseases, 2014; 6(1): e2014070.
14. Rom W, Garay SM. Tuberculosis. New York: Wolters Kluwer; 2004.
15. Cantwell MF, Snider DE, Cauthen GM, Onorato IM. Epidemiology of tuberculosis in the United States, 1985 through 1992. JAMA, 1994; 272(7): 535-539.
16. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2015. Atlanta: Centers for Disease Control and Prevention, 2016.
17. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 1997. Atlanta: Centers for Disease Control and Prevention, 1998.
18. Auld SC, Click ES, Heilig CM, Miramontes R, Cain KP, Bisson GP, Mac Kenzie WR. Tuberculin skin test result and risk of death among persons with active TB. PLoS ONE, 2013; 8(11): e78779.
19. Auld SC, Click ES, Heilig CM, Miramontes R, Cain KP, Bisson GP, Mac Kenzie WR. Association between tuberculin skin test result and clinical presentation of tuberculosis disease. BMC Infectious Diseases, 2013; 13:460.

20. Delgado JC, Tsai EY, Thim S, Baena A, Boussiotis VA, Reynes J-M, Sath S, Grojean P, Yunis EJ, Goldfeld AE. Antigen-specific and persistent tuberculin anergy in a cohort of pulmonary tuberculosis patients from rural Cambodia. Proceedings of the National Academy of Sciences of the United States of America, 2002; 99(11): 7576-7581.
21. Boussiotis VA, Tsai EY, Yunis EJ, Thim S, Delgado JC, Dascher CC, Berezovskaya A, ROusset D, Reynes J-M, Goldfeld AE. IL-10 producing T cells suppress immune response in anergic tuberculosis patients. The Journal of Clinical Investigation, 2000; 105(9): 1317-1325.
22. Kurbatova EV, Cegielski JP, Lienhardt C, Rattanawadee A, Bayona J, Becerra MC, Caoili J, Contreras C, Dalton T, Danilovis M, Demikhova OV, Ershova J, Gammino VM, Gelmanova I, Heilig CM, Jou R, Kazennyy B, Keshavjee S, Kim HJ, Kliiman K, Kvasnovsky C, Leimane V, Mitnick CD, Quelapio I, Riekstina V, Smith SE, Tupasi T, van der Walt M, Vasilyeva IA, Via LE, Viiklepp P, Volchenkov G, Walker AT, Wolfgang M, Yagui M, Zignol M. Sputum culture conversion as a prognostic marker for end-of-treatment outcome in patients with multidrug-resistant tuberculosis: a secondary analysis of data from two observational cohort studies. Lancet Respiratory Medicine, 2015; 3: 201-209.
23. Holtz TH, Sternberg M, Kammer S, Laserson KF, Riekstina V, Zrovska E, Skripconoka V, Wells CD, Leimane V. Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. Annals of Internal Medicine, 2006; 144(9): 650-659.
24. American Thoracic Society. Treatment of tuberculosis and tuberculosis infection in adults and children. American Journal of Respiratory and Critical Care Medicine, 1994; 149(5): 1359-1374.
25. Kurbatova EV, Gammino VM, Bayona J, Becerra MC, Danilovitz M, Falzon D, Gelmanova I, Keshavjee S, Leimane V, Mitnick CD, Quelapio MI, Riekstina V, Taylor A, Viiklepp P, Zignol M, Cegielski JP. Predictors of sputum culture conversion among patients treated for multidrug-resistant tuberculosis. The International Journal of Tuberculosis and Lung Disease, 2012; 16(10):1335-1343.

26. Liu Z, Shilkret KL, Ellis HM. Predictors of sputum culture conversion among patients with tuberculosis in the era of tuberculosis resurgence. Archives of Internal Medicine, 1999;
27. Cegielski JP, Arab L, Cornoni-Huntley J. Nutritional risk factors for tuberculosis among adults in the United States, 1971-1992. American Journal of Epidemiology, 2012; 176(5): 409-422.
28. Koethe JR, von Reyn CF. Protein-calorie malnutrition, macronutrient supplements, and tuberculosis. The International Journal of Tuberculosis and Lung Disease, 2016; 20(7): 857-863.
29. Cegielski JP, McMurray DN. The relationship between malnutrition and tuberculosis: evidence from studies in humans and experimental animals. The International Journal of Tuberculosis and Lung Disease, 2004; 8(3): 286-298.
30. Cegielski JP, McMurray DN. Tuberculosis: Nutritional Management. Encyclopedia of Human Nutrition, 2013; 3(4): 293-298.

TABLES

Table 1: Unweighted and weighted[a] frequencies of sociodemographic and clinical characteristics of hospitalized TB patients in UTHSC at Tyler (1985-2010)

Characteristic		Unweighted n (%)	Weighted n (%)
Total		**491 (100)**	**1,333 (100)**
Tuberculin skin test[b]			
	Negative	88 (17.9)	228 (17.1)
	Positive	403 (82.1)	1,105 (82.9)
Culture conversion[c]			
	Converted	278 (56.6)	700 (52.5)
	Not converted	213 (43.4)	633 (47.5)
Age (years)			
	0-24	55 (11.2)	172 (12.9)
	25-44	180 (36.7)	454 (34.1)
	45-64	186 (37.9)	488 (36.6)
	65+	69 (14.1)	215 (16.1)
Sex			
	Female	118 (24.0)	337 (25.3)
	Male	373 (76.0)	996 (74.7)
Race			
	Asian	18 (3.7)	39 (2.9)
	Black	170 (34.6)	467 (35.0)
	Hispanic	118 (24.0)	293 (22.0)
	White	174 (35.4)	508 (38.1)
	Other or unknown	11 (2.2)	26 (2.0)
Birthplace[d]			
	Foreign-born	175 (35.6)	444 (33.3)
	US-born	314 (64.0)	884 (66.3)
IV drug use			
	Yes	35 (7.1)	83 (6.2)
	No	434 (88.4)	1,187 (89.0)
Illegal drug use			
	Yes	76 (15.5)	158 (11.9)
	No	397 (80.9)	1,121 (84.1)
Alcohol			
	Current drinker	187 (38.1)	517 (38.8)
	Former drinker	72 (14.7)	185 (13.9)
	Never	221 (45.0)	600 (45.0)
Tobacco			
	Current smoker	248 (50.5)	692 (51.9)
	Former smoker	88 (17.9)	212 (15.9)
	Never	146 (29.7)	407 (30.5)
Homeless			
	Yes	95 (19.4)	277 (20.8)
	No	371 (75.6)	1,002 (75.2)
Health insurance			
	Yes	12 (2.4)	37 (2.8)
	No	466 (94.9)	1,264 (94.8)

Table 1 *Continued*

Characteristic		Unweighted n (%)	Weighted n (%)
HIV status			
	Positive	30 (6.1)	30 (2.3)
	Negative	445 (90.6)	1,257 (94.3)
Malnutrition			
	Yes	132 (26.9)	368 (27.6)
	No	346 (70.5)	935 (70.1)
BMI			
	Underweight (<18.5)	85 (17.3)	230 (17.3)
	Normal (18.5-24.9)	131 (26.7)	354 (26.6)
	Overweight or obese (≥25)	29 (5.9)	81 (6.1)
AFB smear test			
	Positive	399 (81.3)	1,047 (78.5)
	Negative	77 (15.7)	244 (18.3)
TB Treatment[e]			
First line drugs			
	Yes	458 (93.3)	1,257 (94.3)
	No	33 (6.7)	76 (5.7)
Second line drugs			
	Yes	213 (43.4)	468 (35.1)
	No	278 (56.6)	865 (64.9)
MDR-TB[f]			
	Yes	47 (9.6)	47 (3.5)
	No	365 (74.3)	1,054 (79.1)
Cancer			
	Yes	183 (37.3)	545 (40.9)
	No	288 (58.7)	736 (55.2)
Diabetes mellitus			
	Yes	66 (13.4)	156 (11.7)
	No	414 (84.3)	1,151 (86.3)
History of TB			
	Yes	121 (24.6)	231 (17.3)
	No	370 (75.4)	1,102 (82.7)

[a]Weighted frequencies reflect how the study population was selected: 25% simple random sample of all TB patients and a 100% sample of individuals who were HIV-positive, multidrug-resistant, and whose serum drug levels were measured

[b]Any tuberculin skin test result available from a different healthcare provider in the past or from the patient's first visit to the study site

[c]Three consecutive negative sputum cultures required to be considered to have conversion

[d]Those with unknown origin were considered foreign-born if the records indicated that they received BCG vaccination or immigrated to the U.S.

[e]Took respective drugs for at least 30 days

[f]Defined as resistant to at least isoniazid and rifampin to be considered as multi-drug resistant (MDR)

TB= tuberculosis; IV = intravenous; HIV = human immunodeficiency virus; BMI = body mass index; AFB = acid-fast bacilli; MDR= multidrug-resistant

Table 2: Comparison of weighted[a] frequencies of sociodemographic and clinical characteristics of hospitalized TB patients in UTHSC at Tyler (1985-2010) by tuberculin skin test (TST)[b] result

		TST Positive n (%)	TST Negative n (%)	X^2	P-value[c]
Total		**1,105 (82.9)**	**228 (17.1)**	**--**	**--**
Age (years)				11.1	**0.02**
	≤24	165 (95.9)	7 (4.1)		
	25-44	383 (84.4)	71 (15.6)		
	45-64	385 (78.9)	103 (21.1)		
	65+	168 (78.1)	47 (21.9)		
Sex				2.2	0.20
	Female	294 (87.2)	43 (12.8)		
	Male	811 (81.4)	185 (18.6)		
Race				1.9	0.21
	Non-white	675 (84.5)	124 (15.5)		
	White	404 (79.5)	104 (20.5)		
Birthplace[d]				0.1	0.78
	Foreign-born	371 (83.6)	73 (16.4)		
	US-born	729 (82.5)	155(17.5)		
Illegal drug use				0.6	0.40
	Yes	137 (86.7)	21 (13.3)		
	No	926 (82.6)	195 (17.4)		
Alcohol				1.1	0.64
	Current drinker	431 (83.4)	86 (16.6)		
	Former drinker	161 (87.0)	24 (13.0)		
	Never	490 (81.7)	110 (18.3)		
Tobacco				2.4	0.39
	Current smoker	581 (84.0)	111 (16.0)		
	Former smoker	164 (77.4)	48 (22.6)		
	Never	346 (85.0)	61 (15.0)		
Homelessness				0.1	0.79
	Yes	226 (81.6)	51 (18.4)		
	No	830 (82.8)	172 (17.2)		
HIV status				4.1	**0.002**
	Positive	18 (60.0)	12 (40.0)		
	Negative	1,047 (83.3)	210 (16.7)		
Malnutrition				7.7	**0.01**
	Yes	277 (75.3)	91 (24.7)		
	No	803 (85.9)	132 (14.1)		
AFB smear test				12.1	**0.001**
	Positive	840 (80.2)	207 (19.8)		
	Negative	233 (95.5)	11 (4.5)		

Table 2 *Continued*

		TST Positive n (%)	TST Negative n (%)	X^2	P-value[c]
MDR-TB[e]				0.1	0.61
	Yes	40 (85.1)	7 (14.9)		
	No	865 (82.1)	189 (17.9)		
TB treatment[f]					
First line drugs				0.3	0.61
	Yes	1,039 (82.7)	218 (17.3)		
	No	66 (86.8)	10 (13.2)		
Second line drugs				7.2	**0.02**
	Yes	359 (76.7)	109 (23.3)		
	No	746 (86.2)	119 (13.8)		

[a]Weighted frequencies reflect how the study population was selected: 25% simple random sample of all TB patients and a 100% sample of individuals who were HIV-positive, multidrug-resistant, and whose serum drug levels were measured

[b]Any TST available from a different healthcare provider in the past or from the patient's first visit to the study site

[c]The Pearson's chi-square test

[d]Those with unknown origin were considered foreign-born if the records indicated that they received BCG vaccination or immigrated to the U.S.

[e]Defined as resistant to at least isoniazid and rifampin to be considered as multi-drug resistant (MDR)

[f]Took respective drugs for at least 30 days

TB = tuberculosis; HIV = human immunodeficiency virus; AFB = acid-fast bacilli; MDR = multidrug-resistant

Table 3: Bivariate analysis of tuberculin skin test (TST)[a] as the main exposure variable and risk factors for TB in relation to sputum culture conversion[b] among hospitalized TB patients in UTHSC at Tyler (1985-2010)

		Converters[c] n (%)	Non-Converters n (%)	X^2	P-value[d]
Total		**700 (52.5)**	**633 (47.5)**	--	--
TST				1.3	0.31
	Negative	107 (46.9)	121 (53.1)		
	Positive	593 (53.7)	512 (46.3)		
Age (years)				26.3	**0.0001**
	≤24	50 (29.1)	122 (70.9)		
	25-44	252 (55.5)	202 (44.5)		
	45-64	308 (63.1)	180 (36.9)		
	65+	90 (41.9)	125 (58.1)		
Sex				31.2	**<0.0001**
	Female	104 (30.9)	400 (69.1)		
	Male	596 (59.8)	233 (40.2)		
Race					
	Non-white	461 (57.7)	338 (42.3)	7.5	**0.02**
	White	228 (44.9)	280 (55.1)		
Birthplace[d]				0.6	0.47
	Foreign-born	221 (49.8)	223 (50.2)		
	US-born	474 (53.6)	410 (46.4)		
Illegal drug use				12.3	**0.0013**
	Yes	117 (74.1)	41 (25.9)		
	No	556 (49.6)	565 (50.4)		
Alcohol				33.4	**<0.0001**
	Current drinker	327 (63.2)	190 (36.8)		
	Former drinker	127 (68.6)	58 (31.4)		
	Never	231 (38.5)	369 (61.5)		
Tobacco				24.2	**<0.0001**
	Current smoker	411 (59.4)	281 (40.6)		
	Former smoker	133 (62.7)	79 (37.3)		
	Never	147 (36.1)	260 (63.9)		
Homelessness				50.5	**<0.0001**
	Yes	230 (83.0)	47 (17.0)		
	No	432 (43.1)	570 (56.9)		
HIV status				0.9	0.13
	Positive	20 (66.7)	10 (33.3)		
	Negative	658 (52.3)	599 (47.7)		
Malnutrition				17.2	**0.0002**
	Yes	251 (68.2)	117 (31.8)		
	No	441 (47.2)	494 (52.8)		
AFB smear test				7.4	**0.02**
	Positive	595 (56.8)	452 (43.2)		
	Negative	100 (41.0)	144 (59.0)		
MDR-TB[e]				3.9	**0.002**
	Yes	38 (80.9)	9 (19.1)		
	No	602 (57.1)	452 (42.9)		

Table 3 *Continued*

		Converters[c] n (%)	Non-Converters n (%)	X^2	P-value[d]
TB treatment[f]					
First line drugs				5.2	**0.03**
	Yes	676 (53.8)	581 (46.2)		
	No	24 (31.6)	52 (68.4)		
Second line drugs				8.3	**0.009**
	Yes	287 (61.3)	181 (38.7)		
	No	413 (47.7)	452 (52.3)		

[a]Any TST available from a different healthcare provider in the past or from the patient's first visit to the study site
[b]Three consecutive negative sputum cultures required to be considered to have conversion
[c]The frequencies of converters and non-converters (or censored) were weighted to reflect how the study population was selected: 25% simple random sample of all TB patients and a 100% sample of individuals who were HIV-positive, multidrug-resistant, and whose serum drug levels were measured
[d]The Pearson Chi-square test was performed to measure the association between categorical predictors and the dichotomized culture conversion outcome
[d]Those with unknown origin were considered foreign-born if the records indicated that they received BCG vaccination or immigrated to the U.S.
[e]Defined as resistant to at least isoniazid and rifampin to be considered as multi-drug resistant (MDR)
[f]Took respective drugs for at least 30 days
TB = tuberculosis; HIV = human immunodeficiency virus; AFB = acid-fast bacilli; MDR = multidrug-resistant

Table 4: Unadjusted hazard ratios[a] of time to sputum culture conversion[b] by risk factors among hospitalized TB patients[c] in UTHSC at Tyler (1985-2010)

		Crude HR	95%CI
TST[d]			
	Negative	0.88	(0.61-1.26)
	Positive	REF	--
Age (years)			
	≤24	0.72	(0.40-1.28)
	25-44	REF	--
	45-64	1.30	(0.97-1.74)
	65+	1.32	(0.84-2.09)
Sex			
	Female	0.70	(0.46-1.02)
	Male	REF	--
Race			
	Non-white	1.13	(0.86-1.50)
	White	REF	--
Birthplace[e]			
	Foreign-born	*	*
	US-born	REF	--
Illegal drug use			
	Yes	1.07	(0.76-1.50)
	No	REF	--
Alcohol			
	Current drinker	1.17	(0.85-1.62)
	Former drinker	1.36	(0.94-1.97)
	Never	REF	--
Tobacco			
	Current smoker	1.16	(0.83-1.63)
	Former smoker	1.33	(0.87-2.02)
	Never	REF	--
Homelessness			
	Yes	1.24	(0.92-1.66)
	No	REF	--
HIV status			
	Positive	§	§
	Negative	REF	--
Malnutrition			
	Yes	1.26	(0.96-1.65)
	No	REF	--
AFB smear test			
	Positive	0.50	**(0.29-0.85)**
	Negative	REF	--
MDR-TB[f]			
	Yes	§	§
	No	REF	--

Table 4 *Continued*

		Crude HR	95%CI
TB treatment[g]			
First line drugs			
	Yes	1.72	(0.79-3.72)
	No	REF	--
Second line drugs			
	Yes	1.01	(0.78-1.31)
	No	REF	--

[a]HR <1 means that a patient with a risk factor, on average, has lower chances of converting sputum culture than a patient without this risk factor

[b]Three consecutive negative sputum cultures required to be considered to have conversion

[c]Study population weighted by selection probabilities: 25% simple random sample of all TB patients except for a 100% sample of HIV-positive patients, drug-resistant TB patients, and those whose drug serum levels were measured

[d]Any TST available from a different healthcare provider in the past or from the patient's first visit to the study site

[e]Those with unknown origin were considered foreign-born if the records indicated that they received BCG vaccination or immigrated to the U.S.

*Foreign birth variable violated the Proportional Hazards assumption and hence was not included in a regular Cox Proportional Hazard model

[f]Defined as resistant to at least isoniazid and rifampin to be considered as multi-drug resistant (MDR)

[g]Took respective drugs for at least 30 days

§No HR obtained because that risk factor was one of the stratification criteria for 100%-sampled patients and hence was not added to the Cox Proportional Hazard model using 'Proc Surveyphreg' in SAS (version 9.3)

TB = tuberculosis; TST = tuberculin skin test; HR= hazard ratio; HIV = human immunodeficiency virus; AFB = acid-fast bacilli; MDR = multidrug-resistant

Table 5: Adjusted hazard ratios[a] of tuberculin skin test (TST)[b] and initial sputum culture conversion[c] by nutritional status among TB patients hospitalized[d] in UTHSC Tyler (1985-2010)

	Malnourished		Adequately Nourished	
	aHR	**95% CI**	**aHR**	**95% CI**
TST				
Negative	0.58	(0.31-1.07)	1.39	(0.87-2.22)
Positive	REF	--	REF	--

[a]aHR <1 means that when adjusted for sex, race, AFB smear test, and TB treatment, a patient with a negative TST, on average, has lower chances of converting sputum culture compared to a positive-TST patient

[b]Any TST available from a different healthcare provider in the past or from the patient's first visit to the study site

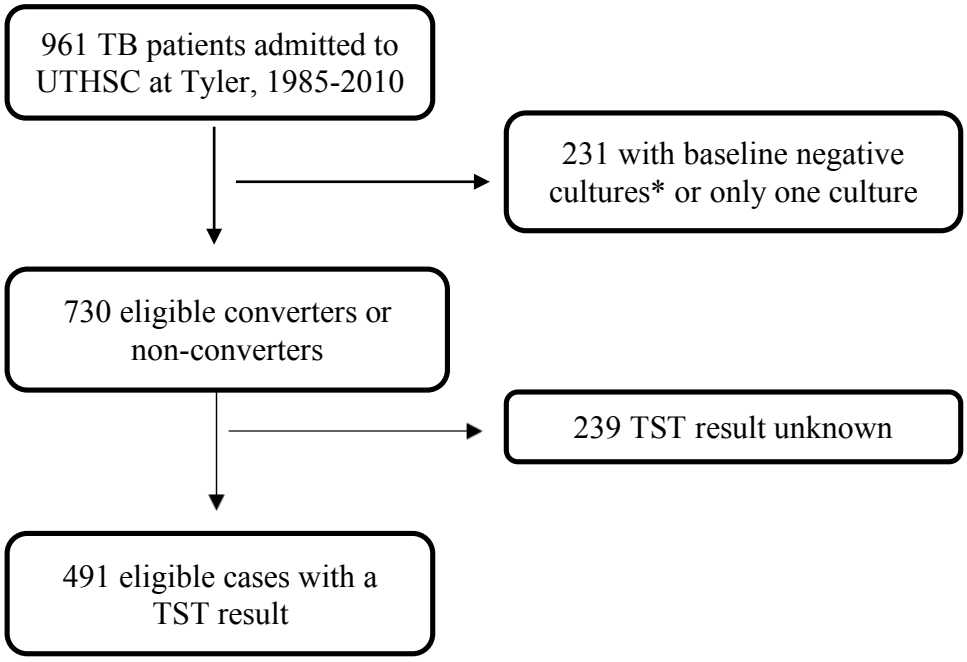
[c]Three consecutive negative sputum cultures required to be considered to have conversion

[d]The hospitalized cohort weighted by selection probabilities: 25% simple random sample of all TB patients except for a 100% sample of HIV-positive patients, drug-resistant TB patients, and those whose drug serum levels were measured

aHR = adjusted hazard ratio; CI = confidence interval

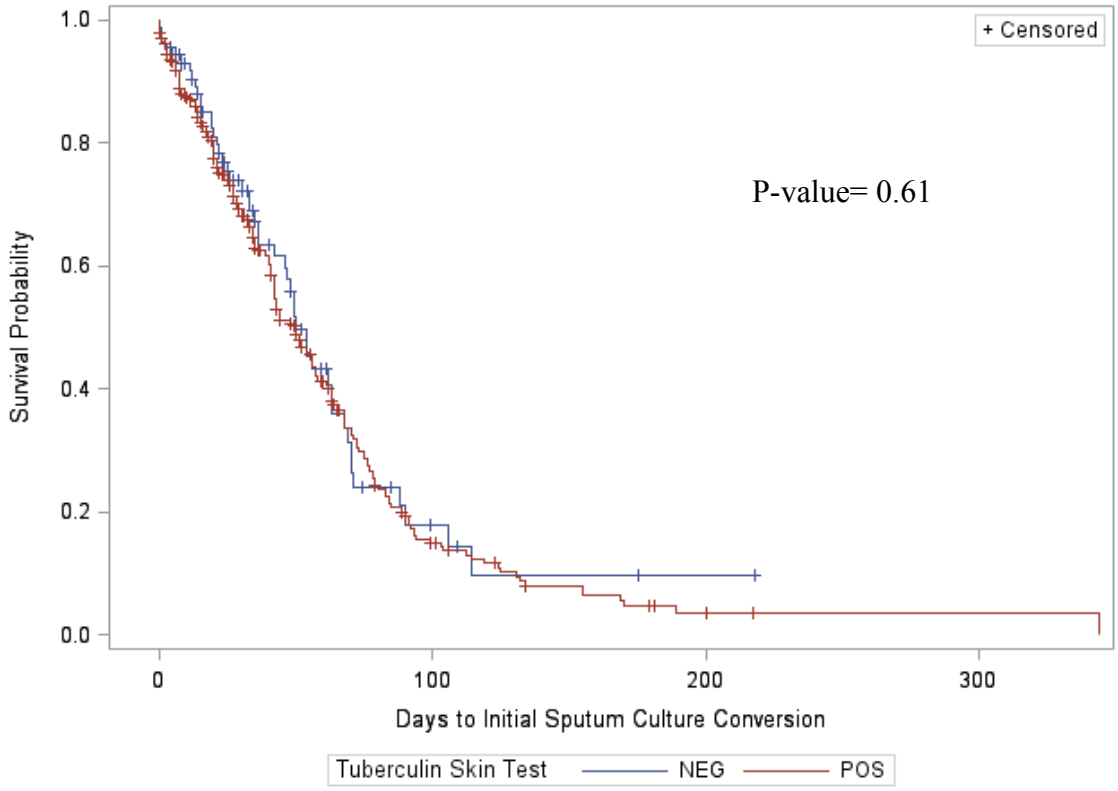
FIGURES

Figure 1: Selection of TB patients admitted to UTHSC at Tyler during 1985 through 2010 for inclusion in the analysis of the relationship between tuberculin skin test (TST) results and sputum culture conversion



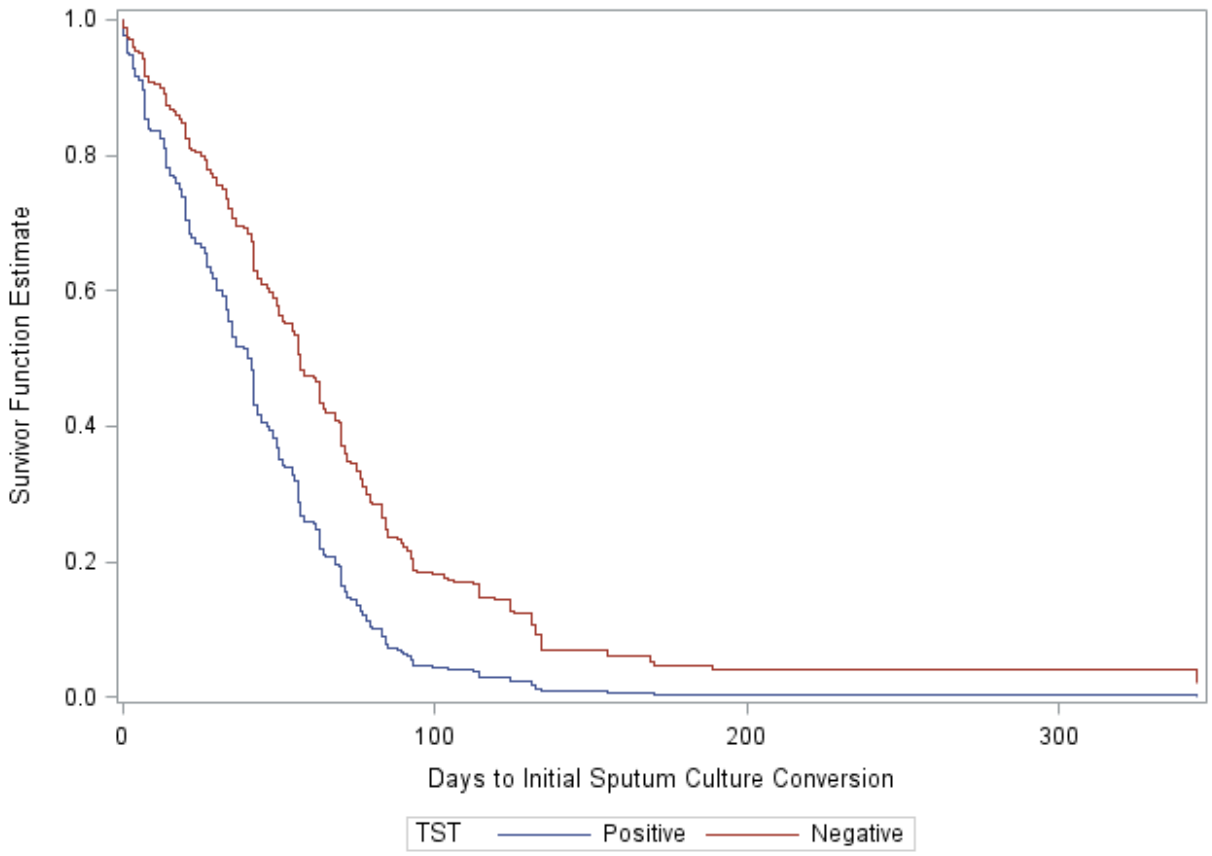
*The concept of "convert to negative" assumes that individuals start out with a positive sputum culture. Therefore, those who had only two or three cultures in total were considered to have conversion if the first culture was positive, and the last culture was negative.

Figure 2: Kaplan-Meier survival curve of time to initial sputum culture conversion versus tuberculin skin test (TST) results in 491 TB patients in UTHSC at Tyler (1985-2010)



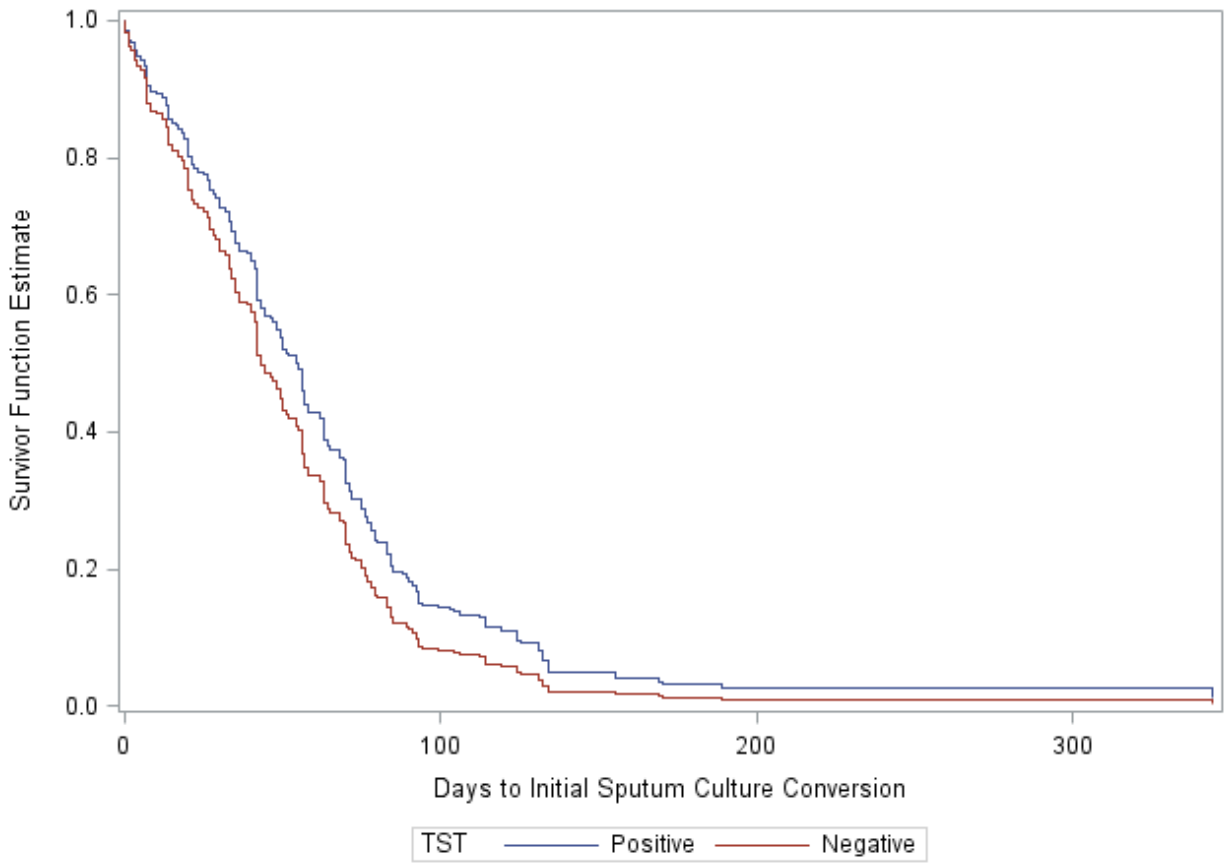
The p-value reflects the results of the log-rank test of the equality of the two survival curves

Figure 3: Adjusted* survival curve of time to initial sputum culture conversion versus tuberculin skin test (TST) results in malnourished TB patients hospitalized in UTHSC at Tyler (1985-2010)



*Adjusted for mean values of covariates included in the Cox proportional hazard model: age, sex, race, acid-fast bacilli smear status, first line drugs, and second line drugs

Figure 4: Adjusted* survival curve of time to initial sputum culture conversion versus tuberculin skin test (TST) results in nourished TB patients hospitalized in UTHSC at Tyler (1985-2010)



*Adjusted for mean values of covariates included in the Cox proportional hazard model: age, sex, race, acid-fast bacilli smear status, first line drugs, and second line drugs

CHAPTER III: PUBLIC HEALTH IMPLICATIONS AND FUTURE DIRECTIONS

Since its introduction in the late nineteenth century, the tuberculin skin test (TST) has been widely used to detect infection with *Mycobacterium tuberculosis*. Despite its popularity, TST is an imperfect screening method for persons with TB infection or disease. Some people with active tuberculosis (TB) do not react to the skin test's antigen as they should and subsequently produce a negative instead of positive TST result. This false-negative TST result could mislead clinicians and hinder individuals with latent TB infection from receiving treatment on time as to prevent progression to TB disease. It is estimated that about 25% of active TB patients might be anergic or have a negative TST result. Previous studies have speculated that some immunological caveats might be associated with a negative TST result and unfavorable outcomes including death and worse disease manifestations.

Overall, our study introduces the relationship between TST result, a marker of host immune response, and time to initial sputum culture conversion as an indicator of responsiveness to TB treatment. We encourage future investigators to repeat this analysis but consider the following: the delay in TB treatment, the timing of the skin test relative to the start of treatment, the size of the TST reaction rather than treating it as dichotomous, and the number of days to conversion from the date of treatment initiation rather than from the date of admission.

The differences in the effect of TST result on sputum culture conversion time by nutritional status, although not statistically significant, suggest a possible biological mechanism where malnutrition weakens CMI, which then leads to less vigorous killing of *M. tuberculosis*. Future study of the association between nutritional status as the main exposure and sputum culture conversion as the outcome may help identify strategies to improve clinical care for persons with active TB, especially if they are malnourished. More incorporation of tuberculin skin testing in clinical trials and prospective studies is recommended to better understand both the host immune system and the mycobacterial pathogen.

APPENDIX: Statistical Analysis System (Version 9.3) Code

```
************************************************************;
* Texas TB Data- TST and Culture Conversion Study           ;
* Master of Public Health Thesis                            ;
* Student: Jennifer Kim, BS                                 ;
* Field Advisor: Peter Cegielski, MD, MPH                   ;
* Faculty Advisor: Kenneth Castro, MD, FIDSA                ;
* Department of Epidemiology                               ;
* Rollins School of Public Health                          ;
* Emory University                                         ;
* November 2016 - April 2017                               ;
* Purpose: to compare time to sputum culture conversion     ;
* between negative-TST and positive-TST patients using      ;
* clinical TB data from Texas, U.S.                         ;
************************************************************;

libname txtb 'h:\thesis\dataset';


*********************************************************;
*Univariate analysis: clean up or create new variables   ;
*as necessary, compute basic frequencies and statistics, ;
*both unweighted and weighted                            ;
*********************************************************;


*-----------------------;
*Hospital admission data;
*-----------------------;

/*exclusions*/
data abcd;
      set txtb.abcd_all_updated;

      **the following are the 262 individuals with baseline negative
      cultures, hence could not convert, and there were 38 individuals with
      only one culture, but the concept of negative conversion implies
      that they should have had at least two cultures;
      **patients with unknown TST were also excluded;

      if ptindex in (*) then DELETE;
      if ptid in (*) then DELETE;
run;

/*the number of unique individuals in the analysis*/
proc freq data=abcd noprint;
      tables ptid/out=numid;
run;
proc freq data=numid;
      table count;
run;
```

```
/*sort by unique patient id to restrict analysis to each individual*/
proc sort data=abcd;
      by ptid;
run;

/*make a new temporary admissions dataset that sorts patients by their first
hospital admission*/
data admission;
      set abcd;
      by ptid;
      if first.ptid;


            **make a new variable called 'foreign' for us vs. foreign-born
            individuals using 'birthplace';
            length foreign $7;

            foreign='NO';*everyone is U.S.-born individuals unless fall into
            the following exclusions below;

            if find(birthplace, 'MX', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'MEX', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'MAXI', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'MONTERREY', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'MATAMOROS', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'RIO', 'i') ge 1 then foreign='YES';

            if find(birthplace, 'INDIA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'Burma', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'CD', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'COLOM', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'CHIHUAHUA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'CHUNTHABURI', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'Cambodia', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'Cuba', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'DELIA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'ETHIOPIA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'CHINA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'HONDURAS', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'VIET', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'IRAPUTO', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'IRAN', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'KOREA', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'ZR', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'MUZQUIZ', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'PAKISTAN', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'PHILLIPINES', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'PLAYVICENTE', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'PUERTO', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'HAITI', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'SAIGON', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'SAIN', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'SALVADOR', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'SAN JUAN', 'i') ge 1 then foreign='YES';
            if find(birthplace, 'SUDAN', 'i') ge 1 then foreign='YES';
```

```
if birthplace='HIDALGO' then foreign='YES';
if birthplace='NEPAL' then foreign='YES';
if birthplace='Nigeria' then foreign='YES';
if birthplace='El Salvador' then foreign='YES';
if birthplace='England' then foreign='YES';
if birthplace='ITALY' then foreign='YES';
if birthplace='KENYA' then foreign='YES';
if birthplace='PINGTUNG' then foreign='YES';
if birthplace='Peru' then foreign='YES';
if birthplace='SAN CIRO DEACOSTA, FN' then foreign='YES';
if birthplace='SAN LUIS POTOSI' then foreign='YES';

if birthplace in (" ", "UNK", "UNKNOWN", "UNKNOWN AR", "unknown")
then foreign=' '; *missing or unknown birthplace;


**create a new bcg variable that combines 'Y' and 'P' as yes bcg;
if bcg in ('Y','P') then bcg1='YES';
else bcg1='NO';
if foreign=' ' and bcg1='YES' then foreign='YES';


**create a new skin test variable to combine prior and at
admission results;
if pst1res in ('N',' ') and pst2res='N' then tstpri='NEG'; *if
either skin test is negative, code it as negative result;
if pst2res in ('N',' ') and pst1res='N' then tstpri='NEG';
if 0<=pst1siz1<5 or 0<=pst2siz1<5 then tstpri='NEG'; *if
transverse diameter less than 5mm, negative result;
if pst1res in ('P','Y') or pst2res in ('P','Y') then
tstpri='POS'; *if either skin test is positive, code it as
positive result;
if pst1siz1 ge 5 or pst2siz1 ge 5 then tstpri='POS'; *if
transverse diameter greater than or equal to 5mm, positive
result;

if 0<=ppd1sz1<5 then tstadm='NEG';*transverse diameter less than
5mm, negative tst;
if ppd1sz1 ge 5 then tstadm='POS';*transverse diameter greater
than or equal to 5mm, positive tst;
if 0<=ppd2sz1<5 then tstadm='NEG';
if ppd2sz1 ge 5 then tstadm='POS';

if tstadm='POS' or tstpri='POS' then anytst='POS';
if tstadm in ('NEG',' ') and tstpri='NEG' then anytst='NEG';
if tstpri in ('NEG',' ') and tstadm='NEG' then anytst='NEG';


**create a bmi variable;
if 4<=htin<=7 then htin=12*floor(htin)+(10*(htin-floor(htin)));
if ptindex=1048.01 then awtkgs=.; *fix individual mistakes in the
data;
if ptindex=1053.01 then do;
            awtlbs=.;
            dwtlbs=.;
            htin=.;
      end;
```

```
if ptindex=1106.01 then do;
            age=.;
            awtkgs=.;
            htin=.;
      end;
if ptindex=1126.01 then htcm=.;
if ptindex=2034.01 then htin=.;
if ptindex=5009.01 then awtkgs=dwtlbs*0.4536-1;

if awtkgs=. then weightkg=awtlbs*0.4536; *calculate weight in kg,
weightkg=awtkgs;
heightm=htcm/100; *calculate height in m;
if htcm=. then heightm=htin*0.0254;
if ptindex=475.01 then heightm=.; *fix 0 height to missing for
this particular individual;

bmi=weightkg/heightm**2; *bmi by definition is weight in kg over
height in meter squared;

if age lt 19 then bmi=.; *bmi not valid for children because
their nutritional status is measured in terms of z-scores;

**create a bmi group variable;
if bmi lt 18.5 then bmigroup=1;
if 18.5<=bmi<25 then bmigroup=2;
if bmi ge 25 then bmigroup=3;
if bmi eq . then bmigroup=.;

**dichotomize the tb history variable (either had tb in the past
or was diagnosed at admission);
if hxtb='OCC' or hxtb='OTH' or hxtb='REC' or hxtb='REL' then
hxtb_yn='YES';
if hxtb='NEW' then hxtb_yn='NO';
if hxtb=' ' then hxtb_yn=' ';

**create dummies for 4 age groups;
if 0<=age<25 then do;
      age0_24=1;
      age45_64=0;
      age65_100=0;
end;
if 25<=age<45 then do;
      age0_24=0;
      age45_64=0;
      age65_100=0;
end;
if 45<=age<65 then do;
      age0_24=0;
      age45_64=1;
      age65_100=0;
end;
if 65<=age<101 then do;
      age0_24=0;
      age45_64=0;
      age65_100=1;
end;
```

```
if age eq . then do;
      age0_24=.;
      age45_64=.;
      age65_100=.;
end;

**create another age group variable for crude analysis;
if 0<=age<25 then agegroup=1;
if 25<=age<45 then agegroup=2;
if 45<=age<65 then agegroup=3;
if age ge 65 then agegroup=4;
if age eq . then agegroup=.;

**combine unknowns and missings into one category for
sociodemographic variables;
if race='O' or race='X' or race='U' or race=' ' then
race_group='';
if race='A' then race_group='A';
if race='B' then race_group='B';
if race='H' then race_group='H';
if race='W' then race_group='W';

if healcare=' ' or healcare='U' then insurance=' ';
if healcare='Y' then insurance='Y';
if healcare='N' then insurance='N';

if alcohol=' ' or alcohol='U' then alcohol_use=' ';
if alcohol='C' or alcohol='Y' then alcohol_use='C';
if alcohol='N' then alcohol_use='N';
if alcohol='F' then alcohol_use='F';

if tobac=' ' or tobac='U' then tobac_use=' ';
if tobac='C' or tobac='Y' then tobac_use='C';
if tobac='N' then tobac_use='N';
if tobac='F' then tobac_use='F';

if homeless=' ' or homeless='U' then homeless_status=' ';
if homeless='N' then homeless_status='N';
if homeless='Y' then homeless_status='Y';

if illdrug=' ' or illdrug='U' then illdrug_use=' ';
if illdrug='N' then illdrug_use='N';
if illdrug='Y' then illdrug_use='Y';

if ivdrug=' ' or ivdrug='U' then ivdrug_use=' ';
if ivdrug='N' then ivdrug_use='N';
if ivdrug='Y' then ivdrug_use='Y';

**combine unknowns and missings as into one category for clinical
variables;
if hiv='U' or hiv=' ' then hiv_status=' ';
if hiv='N' then hiv_status='N';
if hiv='Y' then hiv_status='Y';

**create a binary race variable (either white or non-white);
if race_group='W' then white_n=0; *0 if white, 1 otherwise;
if race_group='A' then white_n=1;
```

```
            if race_group='B' then white_n=1;
            if race_group='H' then white_n=1;
            if race_group=' ' then white_n=.;

            **make corrections to date of discharge and date of admission;
            IF PTINDEX= 508.01 THEN DO; DATEADMIT=MDY(4,2,1997);
            DATEDC=MDY(4,16,1997); END;
            IF PTINDEX=1008.01 THEN DO; DATEADMIT=MDY(4,8,1985);
            DATEDC=MDY(4,7,1986);  END;
            IF PTINDEX=1040.01 THEN DO; DATEADMIT=MDY(5,26,1986);
            DATEDC=MDY(6,12,1986); END;
            IF PTINDEX=1107.01 THEN DO; DATEADMIT=MDY(2,5,2003);
            DATEDC=MDY(2,26,2003); END;
            IF PTINDEX=1136.01 THEN DO; DATEADMIT=MDY(1,4,2005);
            DATEDC=MDY(1,9,2005);  END;
            IF PTINDEX=5051.01 THEN DO;
            DATEADMIT=MDY(12,24,2002);DATEDC=MDY(1,22,2003); END;

            **fill in missing values for foreign birth and malnutrition using
            other indicator variables;
            if foreign=' ' and immigrnt='Y' then foreign='YES';
            if foreign=' ' and immigrnt='N' then foreign='NO';
            if foreign=' ' and immigrnt='U' and english='N' then
            foreign='YES';
            if foreign=' ' and immigrnt='U' and english='Y' then
            foreign='NO';

            if malnut=' ' and bmigroup=1 then malnut='YES';
            if malnut=' ' and bmigroup>1 then malnut='NO';

run;


*-----------------------------;
*Drug susceptibility test data;
*-----------------------------;

/*exclusions*/
data dst;
      set txtb.sumdstresults;

      **the following are the 262 individuals with baseline negative
      cultures, hence could not convert, and there were 38 individuals with
      only one culture, but the concept of negative conversion implies
      that they should have had at least two cultures;
      **patients with unknown TST were also excluded;

      if ptindex in (*) then DELETE;
      if ptid in (*) then DELETE;
run;

/*restrict it to unique individual*/
proc sort data=dst;
      by ptid;
run;
```

```
data dst2;
      set dst;
      by ptid;
      if first.ptid;
run;

/*make a new dataset that contains a dichotomous variable for mdr-tb*/
data mdr;
      set dst2;
      if sumdst_inh='R' and sumdst_rif='R' then mdr_tb='Y';
      else if sumdst_inh='S' or sumdst_rif='S' then mdr_tb='N';
      else if sumdst_inh=' ' or sumdst_rif=' ' then mdr_tb=' ';
run;


*-----------------;
*TB treatment data;
*-----------------;

/*exclusions*/
data treatment;
      set txtb.treatment_variables2;

      **the following are the 262 individuals with baseline negative
      cultures, hence could not convert, and there were 38 individuals with
      only one culture, but the concept of negative conversion implies
      that they should have had at least two cultures;
      **patients with unknown TST were also excluded;

      if ptindex in (*) then DELETE;
      if ptid in (*) then DELETE;
run;

/*restrict it to unique individual*/
proc sort data=treatment;
      by ptid;
run;
data treatment2;
      set treatment;
      by ptid;
      if first.ptid;
run;
```

```
*----------------------;
*Mycobacteriology data;
*----------------------;

/*exclusions*/
data myco;
      set txtb.myco_all_ts;

      **the following are the 262 individuals with baseline negative
      cultures, hence could not convert, and there were 38 individuals with
      only one culture, but the concept of negative conversion implies
      that they should have had at least two cultures;
      **patients with unknown TST were also excluded;

      if ptindex in (*) then DELETE;
      if ptid in (*) then DELETE;
run;


/*before creating new datasets for different purposes, sort the above
temporary dataset by ptindex and lab test date*/
proc sort data=myco;
      by ptindex datetblab;
run;

/*create temporary datasets to obtain the frequency of afb smear results*/
data afb (keep=ptindex afbpos);
      set myco;
      by ptindex datetblab;
            if afbsmear not in ('POS','NEG') then DELETE; *afb not done
            altogether uninformative hence omit;
            if first.ptindex then afbpos=0;
            if afbsmear='POS' then afbpos+1; *start counting any positive
            afb;
            if last.ptindex then output;
run;
data afb2 (keep=ptindex ptid afb);
      set afb;
            if afbpos>0 then afb='POS';
            else afb='NEG';

            ptid=int(ptindex);
run;

/*restrict it to unique individual*/
proc sort data=afb2;
      by ptid;
run;
data afb3;
      set afb2;
            by ptid;
            if first.ptid;
run;
```

```
*-------------------------------------------------------;
*Culture conversion data that exclude baseline negative;
*cultures and thosewith only one culture observation   ;
*-------------------------------------------------------;

/*exclusions*/
data conv;
      set txtb.cxconversion_blpos;
      **get unique ptid's; ptid=int(ptindex);

      **the following are patients with unknown tst and were excluded from
      analysis;
            if ptid in (*) then DELETE;
run;

/*restrict it to unique individual*/
proc sort data=conv;
      by ptid;
run;


data cultureconv;
      set conv;
            by ptid;
            if first.ptid;
run;

/*sort other datasets to merge with the culture conversion dataset*/
proc sort data=admission;
      by ptid;
run;
proc sort data=afb3;
      by ptid;
run;
proc sort data=mdr;
      by ptid;
run;
proc sort data=treatment2;
      by ptid;
run;
proc sort data=cultureconv;
      by ptid;
run;

/*merge the culture conversion dataset with abcd, mdr, treatment & afb smear
datasets*/
data final;
      merge cultureconv admission mdr treatment2 afb3;
            by ptid;
      **variable to indicate whether the patient was converted or not;
      if con3first=. then conv=0; *0=censored;
      if con3first=1 then conv=1; *1=the patient had negative culture
      conversion;
      **time to culture conversion or censored;
      if days2con3first=. then days2con3first= datedc-dateadmit; *follow-up
      time= date of discharge - date of admission;
run;
```

```
*****************************************************;
*Univariate analysis- look at one predictor at a time;
*****************************************************;


*------------------------;
*Sociodemographic factors;
*------------------------;

/*categorical variables*/
proc freq data=final;
      tables sex agegroup race_group white_n foreign insurance alcohol_use
tobac_use illdrug_use ivdrug_use homeless_status/missing;
run;
proc surveyfreq data=final missing;
      tables sex agegroup race_group white_n foreign insurance alcohol_use
tobac_use illdrug_use ivdrug_use homeless_status;
            weight smpwt;
            strata stratum;
run;

/*continuous variables*/
proc surveymeans data=final missing sum mean min quartiles max missing;
      var age;
            weight smpwt;
            strata stratum;
run;


*----------------;
*Clinical factors;
*----------------;

/*categorical variables*/
proc freq data=final;
      tables hiv_status malnut diabetes cancer anytst tstpri tstadm bmigroup
hxtb_yn afb conv mdr_tb trtfld7 trtsld7 trtfld14 trtsld14 trtfld30 trtsld30
bcg1/ missing;
run;
proc surveyfreq data=final missing;
      tables hiv_status malnut diabetes cancer anytst tstpri tstadm bmigroup
hxtb_yn afb conv mdr_tb trtfld7 trtsld7 trtfld14 trtsld14 trtfld30 trtsld30
bcg1;
            weight smpwt;
            strata stratum;
run;

/*continuous variables*/
proc univariate data=final;
      var lengthstay days2con3first bmi;
run;

proc surveymeans data=final missing sum mean min quartiles max missing;
      var lengthstay days2con3first bmi;
            weight smpwt;
            strata stratum;
run;
```

```
*************************************************************************;
*Bivariate analysis- crude relationship between covariates and exposure or  ;
outcome                                                                    ;
*************************************************************************;


*--------------------------------------;
*Exposure: tst positive or tst negative ;
*--------------------------------------;

/*chi-square test: sociodemographic factors vs. tuberculin skin test*/

**sex;
proc surveyfreq data=final;
      tables sex*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**age;
proc surveyfreq data=final;
      tables agegroup*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**race;
proc surveyfreq data=final;
      tables white_n*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**foreign-born;
proc surveyfreq data=final;
      tables foreign*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**alcohol use;
proc surveyfreq data=final;
      tables alcohol_use*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**tobacco smoking;
proc surveyfreq data=final;
      tables tobac_use*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**illegal drug use;
proc surveyfreq data=final;
      tables illdrug_use*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**iv drug use;
proc surveyfreq data=final;
      tables ivdrug_use*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**homelessness;
proc surveyfreq data=final;
      tables homeless_status*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

/*clinical factors vs. tuberculin skin test*/

**hiv;
proc surveyfreq data=final;
      tables hiv_status*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**malnutrition;
proc surveyfreq data=final;
      tables malnut*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**diabetes;
proc surveyfreq data=final;
      tables diabetes*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**cancer;
proc surveyfreq data=final;
      tables cancer*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;

**bmi;
proc surveyfreq data=final;
      tables bmigroup*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**afb smear;
proc surveyfreq data=final;
      tables afb*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;
**mdr-tb;
proc surveyfreq data=final;
      tables mdr_tb*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;
**first line drugs (7 days);
proc surveyfreq data=final;
      tables trtfld7*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;

**second line drugs (7 days);
proc surveyfreq data=final;
      table trtsld7*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;

**first line drugs (14 days);
proc surveyfreq data=final;
      tables trtfld14*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;

**second line drugs (14 days);
proc surveyfreq data=final;
      table trtsld14*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;

**first line drugs (30 days);
proc surveyfreq data=final;
      tables trtfld30*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;

**second line drugs (30 days);
proc surveyfreq data=final;
      table trtsld30*anytst/ chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**history of tb;
proc surveyfreq data=final;
      tables hxtb_yn*anytst / chisq row;
            weight smpwt;
            strata stratum;
run;



*-------------------------------------------------------;
*Outcome: had negative culture conversion or censored ;
*-------------------------------------------------------;

/*sociodemographic factors vs. culture conversion*/


**sex;
proc surveyfreq data=final;
      tables sex*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**age;
proc surveyfreq data=final;
      tables agegroup*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**race;
proc surveyfreq data=final;
      tables white_n*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**foreign-born;
proc surveyfreq data=final;
      tables foreign*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**alcohol use;
proc surveyfreq data=final;
      tables alcohol_use*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**tobacco smoking;
proc surveyfreq data=final;
      tables tobac_use*conv / chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**illegal drug use;
proc surveyfreq data=final;
      tables illdrug_use*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**iv drug use;
proc surveyfreq data=final;
      tables ivdrug_use*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**homelessness;
proc surveyfreq data=final;
      tables homeless_status*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

/*clinical factors vs. culture conversion*/

**tst;
proc surveyfreq data=final;
      tables anytst*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**hiv;
proc surveyfreq data=final;
      tables hiv_status*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**malnutrition;
proc surveyfreq data=final;
      tables malnut*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**diabetes;
proc surveyfreq data=final;
      tables diabetes*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**cancer;
proc surveyfreq data=final;
      tables cancer*conv / chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**bmi;
proc surveyfreq data=final;
      tables bmigroup*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**afb smear;
proc surveyfreq data=final;
      tables afb*conv / chisq row;
            weight smpwt;
            strata stratum;
run;
**mdr-tb;
proc surveyfreq data=final;
      tables mdr_tb*conv / chisq row;
            weight smpwt;
            strata stratum;
run;

**first line drugs (7 days);
proc surveyfreq data=final;
      tables trtfld7*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;

**second line drugs (7 days);
proc surveyfreq data=final;
      table trtsld7*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;

**first line drugs (14 days);
proc surveyfreq data=final;
      tables trtfld14*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;

**second line drugs (14 days);
proc surveyfreq data=final;
      table trtsld14*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;

**first line drugs (30 days);
proc surveyfreq data=final;
      tables trtfld30*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;
```

```
**second line drugs (30 days);
proc surveyfreq data=final;
      table trtsld30*conv/ chisq row;
            weight smpwt;
            strata stratum;
run;

**history of tb;
proc surveyfreq data=final;
      tables hxtb_yn*conv / chisq row;
            weight smpwt;
            strata stratum;
run;


*-------------------------------------------------------------------------;
*Screening of covariates (confounding asessment) to decide whether to      ;
*include them in further analysis                                          ;
*-------------------------------------------------------------------------;

/*use proc surveylogistic to compare unadjusted and adjusted for covariates
of interest*/

**create dummies for variables with multiple levels;
data log;
      set final;

            if alcohol_use='C' then alcohol_c=1; else alcohol_c=0;
            if alcohol_use='F' then alcohol_f=1; else alcohol_f=0;
            if alcohol_use='N' then alcohol_c=0; else alcohol_n=0;

            if tobac_use='C' then tobac_c=1; else tobac_c=0;
            if tobac_use='F' then tobac_f=1; else tobac_f=0;
            if tobac_use='N' then tobac_n=1; else tobac_n=0;

run;

**crude association between tst and culture conversion;
proc surveylogistic data=log;
      class anytst(ref='POS');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst;
run;

**adjusted for foreign birth;
proc surveylogistic data=log;
      class anytst(ref='POS') foreign(ref='NO');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst foreign;
run;
```

```
**adjusted for alcohol;
proc surveylogistic data=log;
      class anytst(ref='POS');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst alcohol_c alcohol_f alcohol_n;
run;

**adjusted for tobacco;
proc surveylogistic data=log;
      class anytst(ref='POS');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst tobac_c tobac_f tobac_n;
run;

**adjusted for illegal drug;
proc surveylogistic data=log;
      class anytst(ref='POS') illdrug_use(ref='N');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst illdrug_use;
run;

**adjusted for homelessness;
proc surveylogistic data=log;
      class anytst(ref='POS') homeless_status(ref='N');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst homeless_status;
run;

**adjusted for mdr;
proc surveylogistic data=log;
      class anytst(ref='POS') mdr_tb(ref='N');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst mdr_tb;
run;

**adjusted for history of tb;
proc surveylogistic data=log;
      class anytst(ref='POS') hxtb_yn(ref='NO');
      strata stratum;
      weight smpwt;
      model conv(event='1')= anytst hxtb_yn;
run;
```

```
*************************************************************************;
* Kaplan-meier: generate survival curves with proc lifetest and compare     ;
* rates of conversion between levels of exposure and other factors using    ;
* the log-rank test                                                         ;
*************************************************************************;


*-----------------------;
*Exposure: tst+ vs. tst-;
*-----------------------;

proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata anytst;
      label anytst='Tuberculin Skin Test';
      label days2con3first='Days to Initial Sputum Culture Conversion';
run;


*----------;
*Covariates;
*----------;

**sex;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata sex;
run;

**age;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata age0_24 age45_64 age65_100;
run;

**race;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata white_n; *binary race variable;
run;

**origin of birth;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata foreign;
run;

**illegal drug;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata illdrug_use;
run;
```

```
**hiv;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata hiv_status;
run;

**malnutrition;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata malnut;
run;

**afb smear;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata afb;
run;

**mdr;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata mdr_tb;
run;
**first line drugs;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata trtfld30;
run;

**second line drugs;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata trtsld30;
run;

**bmi;
proc lifetest data=final method=km plots=(lls) plots=survival;
      time days2con3first*conv(0);
      strata bmigroup;
run;
```

```
*************************************************************************;
* Proportional hazards assumption test: graphical approach (compare ln-ln   ;
* curves obtained from above), shoenfeld residuals goodness-of-fit, and     ;
* extended cox model containing time-dependent covariates                   ;
*************************************************************************;


*-----------------------------------------------------------------------;
*From categorical predictors to numeric as necessary for proc phreg;
*-----------------------------------------------------------------------;

data numeric;
      set final;

      if anytst='NEG' then tst=1;
      if anytst='POS' then tst=0;
      if anytst=' ' then tst=.;

      if sex='F' then female=1;
      if sex='M' then female=0;
      if sex=' ' then female=.;

      if foreign='YES' then foreign1=1;
      if foreign='NO' then foreign1=0;
      if foreign=' ' then foreign1=.;

      if illdrug_use='Y' then illdrug_y=1;
      if illdrug_use='N' then illdrug_y=0;
      if illdrug_use=' ' then illdrug_y=.;

      if hiv_status='Y' then hiv_p=1;
      if hiv_status='N' then hiv_p=0;
      if hiv_status=' ' then hiv_p=.;

      if malnut='Y' then malnut1=1;
      if malnut='N' then malnut1=0;
      if malnut=' ' then malnut1=.;
      if afb='POS' then afb_p=1;
      if afb='NEG' then afb_p=0;
      if afb=' ' then afb_p=.;

      if mdr_tb='Y' then mdr_y=1;
      if mdr_tb='N' then mdr_y=0;
      if mdr_tb=' ' then mdr_y=.;

run;
```

```
*-------------------------;
*Schoenfeld residuals (SR);
*-------------------------;

/*tst*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= tst;
      output out=SR_tst ressch=SR_tst;
run;

**the dataset created above that contains the SRs for tst;
data tst_failures;
      set SR_tst;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=tst_failures out=ranked_tst ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_tst nosimple;
      var SR_tst;
      with timerank;
run;

/*sex*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= female;
      output out=SR_sex ressch=SR_sex;
run;

**the dataset created above that contains the SRs for tst;
data sex_failures;
      set SR_sex;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=sex_failures out=ranked_sex ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;
```

```
**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_sex nosimple;
      var SR_sex;
      with timerank;
run;

/*age*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= age;
      output out=SR_age ressch=SR_age;
run;

**the dataset created above that contains the SRs for age;
data age_failures;
      set SR_age;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=age_failures out=ranked_age ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_age nosimple;
      var SR_age;
      with timerank;
run;

/*race*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= white_n;
      output out=SR_race ressch=SR_race;
run;

**the dataset created above that contains the SRs for race;
data race_failures;
      set SR_race;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=race_failures out=ranked_race ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;
```

```
**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_race nosimple;
      var SR_race;
      with timerank;
run;

/*origin of birth*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= foreign1;
      output out=SR_for ressch=SR_for;
run;

**the dataset created above that contains the SRs for origin of birth;
data for_failures;
      set SR_for;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=for_failures out=ranked_for ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_for nosimple;
      var SR_for;
      with timerank;
run;

/*illegal drug*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= illdrug_y;
      output out=SR_illdrug ressch=SR_illdrug;
run;

**the dataset created above that contains the SRs for illegal drug users;
data illdrug_failures;
      set SR_illdrug;
      if conv=1; *only want the failures;
run;
```

```
**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=illdrug_failures out=ranked_illdrug ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_illdrug nosimple;
      var SR_illdrug;
      with timerank;
run;

/*hiv*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= hiv_p;
      output out=SR_hiv ressch=SR_hiv;
run;

**the dataset created above that contains the SRs for hiv;
data hiv_failures;
      set SR_hiv;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=hiv_failures out=ranked_hiv ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_hiv nosimple;
      var SR_hiv;
      with timerank;
run;

/*malnutrition*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= malnut1;
      output out=SR_malnut ressch=SR_malnut;
run;

**the dataset created above that contains the SRs for malnutrition;
data malnut_failures;
      set SR_malnut;
      if conv=1; *only want the failures;
run;
```

```
**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=malnut_failures out=ranked_malnut ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_malnut nosimple;
      var SR_malnut;
      with timerank;
run;

/*afb smear*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= afb_p;
      output out=SR_afb ressch=SR_afb;
run;

**the dataset created above that contains the SRs for afb;
data afb_failures;
      set SR_afb;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=afb_failures out=ranked_afb ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_afb nosimple;
      var SR_afb;
      with timerank;
run;

/*mdr*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= mdr_y;
      output out=SR_mdr ressch=SR_mdr;
run;

**the dataset created above that contains the SRs for mdr;
data mdr_failures;
      set SR_mdr;
      if conv=1; *only want the failures;
run;
```

```
**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=mdr_failures out=ranked_mdr ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_mdr nosimple;
      var SR_mdr;
      with timerank;
run;

/*first line drugs (30 days)*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= trtfld30;
      output out=SR_fld ressch=SR_fld;
run;

**the dataset created above that contains the SRs for fld;
data fld_failures;
      set SR_fld;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=fld_failures out=ranked_fld ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_fld nosimple;
      var SR_fld;
      with timerank;
run;

/*second line drugs (30 days)*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= trtsld30;
      output out=SR_sld ressch=SR_sld;
run;

**the dataset created above that contains the SRs for sld;
data sld_failures;
      set SR_sld;
      if conv=1; *only want the failures;
run;
```

```
**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=sld_failures out=ranked_sld ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_sld nosimple;
      var SR_sld;
      with timerank;
run;

/*bmi*/

**run a cox ph model and calculate SR for each subject;
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= bmi;
      output out=SR_bmi ressch=SR_bmi;
run;

**the dataset created above that contains the SRs for bmi;
data bmi_failures;
      set SR_bmi;
      if conv=1; *only want the failures;
run;

**create a new dataset that ranks the tst_failures dataset by the conversion
time variable;
proc rank data=bmi_failures out=ranked_bmi ties=mean;
      var days2con3first;
      ranks timerank; *the new ranking variable;
run;

**test the correlation between SRs and the ranked failure times;
proc corr data=ranked_bmi nosimple;
      var SR_bmi;
      with timerank;
run;


*------------------------;
*Time-dependent variables;
*------------------------;

/*extended cox model with tst (exposure) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= tst tst_t;
      tst_t= tst*days2con3first;
run;
```

```
/*extended cox model with sex (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= female female_t;
      female_t= female*days2con3first;
run;

/*extended cox model with age (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= age age_t;
      age_t= age*days2con3first;
run;

/*extended cox model with race (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= white_n white_t;
      white_t= white_n*days2con3first;
run;

/*extended cox model with origin of birth (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= foreign1 foreign1_t;
      foreign1_t= foreign1*days2con3first;
run;

/*extended cox model with illegal drug use (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= illdrug_y illdrug_t;
      illdrug_t= illdrug_y*days2con3first;
run;

/*extended cox model with hiv (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= hiv_p hiv_p_t;
      hiv_p_t= hiv_p*days2con3first;
run;

/*extended cox model with malnutrition (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= malnut1 malnut1_t;
      malnut1_t= malnut1*days2con3first;
run;
```

```
/*extended cox model with afb (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= afb_p afb_p_t;
      afb_p_t= afb_p*days2con3first;
run;

/*extended cox model with mdr (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= mdr_y mdr_t;
      mdr_t= mdr_y*days2con3first;
run;

/*extended cox model with first line drugs (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= trtfld30 fld_t;
      fld_t= trtfld30*days2con3first;
run;


/*extended cox model with second line drugs (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= trtsld30 sld_t;
      sld_t= trtsld30*days2con3first;
run;

/*extended cox model with bmi (covariate) x time product term*/
proc surveyphreg data=numeric;
      strata stratum;
      weight smpwt;
      model days2con3first*conv(0)= bmi bmi_t;
      bmi_t= bmi*days2con3first;
run;
```

```
*************************************************************;
* Effect modification between the exposure and covariates ;
*************************************************************;


*-----------------------------------------------------------;
*Likelihood Ratio Test for No-Interaction vs. Interaction;
*-----------------------------------------------------------;

/*create another final dataset that contains interaction terms and transforms
some categorical covariates to numeric to perform phreg*/
data final2;
      set final;

      if anytst='NEG' then tst=1;
      if anytst='POS' then tst=0;
      if anytst=' ' then tst=.;

      if sex='F' then female=1;
      if sex='M' then female=0;
      if sex=' ' then female=.;

      if foreign='YES' then foreign1=1;
      if foreign='NO' then foreign1=0;
      if foreign=' ' then foreign1=.;

      if illdrug_use='Y' then illdrug_y=1;
      if illdrug_use='N' then illdrug_y=0;
      if illdrug_use=' ' then illdrug_y=.;

      if malnut='Y' then malnut1=1;
      if malnut='N' then malnut1=0;
      if malnut=' ' then malnut1=.;

      if afb='POS' then afb_p=1;
      if afb='NEG' then afb_p=0;
      if afb=' ' then afb_p=.;


      **create interaction terms;
      tstxfem= tst*female;
      tstxage= tst*age;
      tstxnwh= tst*white_n;
      tstxfor= tst*foreign1;
      tstxill= tst*illdrug_y;
      tstxmal= tst*malnut1;
      tstxafb= tst*afb_p;
      tstxfld= tst*trtfld30;
      tstxsld= tst*trtsld30;

      mal=malnut1;
run;
```

```
/*The crude association between tst and culture conversion*/
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst/ risklimits;
      weight smpwt;
      strata stratum;
run;

/*With all interaction terms*/
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxfem tstxage tstxill tstxafb tstxfld tstxsld
tstxmal;
      weight smpwt;
      strata stratum;
run;

/*Without interaction terms*/
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1;
      weight smpwt;
      strata stratum;
run;

/*Likelihood ratio test (chunk test): chisquare statistic= -2lnL(red)-
2lnL(full)=48.261 under H0 with df=13*/
data pvalue;
      x= 1- probchi(29.269,7);**obtain p-value;
run;
proc print data=pvalue; *p-value<0.001, reject H0 at alpha 0.05, significant
effect modification;
run;

/*Backward elimination: start with the most insignificant interaction term,
drop one term at a time*/

**drop tstxafb;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxfem tstxage tstxill tstxfld tstxsld tstxmal;
      weight smpwt;
      strata stratum;
run;
**drop tstxafb & tstxsld;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxfem tstxage tstxill tstxfld tstxmal;
      weight smpwt;
      strata stratum;
run;
```

```
**drop tstxafb, tstxsld & tstxilldrug;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxfem tstxage tstxfld tstxmal;
      weight smpwt;
      strata stratum;
run;

**drop tstxafb, tstxsld, tstxilldrug & tstxsex;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxage tstxfld tstxmal;
      weight smpwt;
      strata stratum;
run;

**drop tstxafb, tstxsld, tstxilldrug, tstxsex & tstxfld;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxage tstxmal;
      weight smpwt;
      strata stratum;
run;

*drop tstxafb, tstxsld, tstxilldrug, tstxsex, tstxfld & tstxage;
proc surveyphreg data=final2;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxmal;
      weight smpwt;
      strata stratum;
run;



************************;
* Confounding assessment ;
************************;


*-------------------------------------------------;
*All-possible subsets change in estimate approach;
*-------------------------------------------------;

/*Drop one covariate at a time and compare to the gold standard*/

**gold standard;
proc surveyphreg data=final2;
      domain mal;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= tst female white_n illdrug_y afb_p
trtfld30 trtsld30 malnut1 tstxmal/risklimits;
run;
```

```
**drop illegal drug;
proc surveyphreg data=final2;
      domain mal;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= tst female white_n afb_p trtfld30
trtsld30 malnut1 tstxmal/risklimits;
run;

**drop afb;
proc surveyphreg data=final2;
      domain mal;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= tst female white_n illdrug_y trtfld30
trtsld30 malnut1 tstxmal/risklimits;
run;

**drop illegal drug & afb;
proc surveyphreg data=final2;
      domain mal;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= tst female white_n trtfld30 trtsld30
malnut1 tstxmal/risklimits;
run;



**************************;
* Adjusted survival curves ;
**************************;

/*First find average values of confounders*/
proc means data=final2;
      var female white_n age afb_p trtfld30 trtsld30;
run;

data means;
      input tst female white_n age afb_p trtfld30 trtsld30 malnut1;
      datalines;
      0 0.2403259 0.6375000 45.2306122 0.8382353 0.9327902 0.4338086 0
      1 0.2403259 0.6375000 45.2306122 0.8382353 0.9327902 0.4338086 0
      0 0.2403259 0.6375000 45.2306122 0.8382353 0.9327902 0.4338086 1
      1 0.2403259 0.6375000 45.2306122 0.8382353 0.9327902 0.4338086 1
;
run;

proc phreg data=final2 plots (overlay)=survival;
      model days2con3first*conv(0)= tst white_n age afb_p trtfld30 trtsld30
malnut1 tst|malnut1;
      weight smpwt;
      baseline covariates=means out=adjplot survival=survival/rowid=tst;
run;
```

```
/*Adjusted survival curves*/
proc format;
      value tstf
            1='Negative'
            0='Positive';
run;
proc sgplot data=adjplot;
      title "Adjusted for mean values of covariates among TB patients who
were malnourished";
      step x=days2con3first y=survival/group=tst;
      where malnut1=1;
      format tst tstf.;
      label tst='TST';
      label days2con3first='Days to Initial Sputum Culture Conversion';
run;
title;
proc sgplot data=adjplot;
      title "Adjusted for mean values of covariates among TB Patients who
were not malnourished";
      step x=days2con3first y=survival/group=tst;
      where malnut1=0;
      format tst tstf.;
      label tst='TST';
      label days2con3first='Days to Initial Sputum Culture Conversion';
run;
title;



****************************************;
* Unadjusted hazard ratios of each factor;
****************************************;

**sex;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= female/risklimits;
run;

**age;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= age0_24 age45_64 age65_100/risklimits;
run;

**race;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= white_n/risklimits;
run;
```

```
**foreign birth;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= foreign1/risklimits;
run;

**illegal drug;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= illdrug_y/risklimits;
run;

**alcohol;
proc surveyphreg data=final2;
      class alcohol_use(ref='N');
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= alcohol_use/risklimits;
run;

**tobacco;
proc surveyphreg data=final2;
      class tobac_use(ref='N');
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= tobac_use/risklimits;
run;

**homelessness;
proc surveyphreg data=final2;
      class homeless_status(ref='N');
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= homeless_status/risklimits;
run;

**malnutrition;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= malnut1/risklimits;
run;

**afb smear;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= afb_p/risklimits;
run;
```

```
**fld;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= trtfld30/risklimits;
run;

**sld;
proc surveyphreg data=final2;
      weight smpwt;
      strata stratum;
      model days2con3first*conv(0)= trtsld30/risklimits;
run;
```