Distribution Agreement

In presenting this thesis 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:

Ryan Patrick Kyle	Date

Evaluating Response to Tuberculosis Therapy:
A Generalized Estimating Equation Model

by

Ryan Patrick Kyle

MPH

Epidemiology

Dr. David G. Kleinbaum

Faculty Thesis Advisor

Dr. Charles M. Heilig

Thesis Field Advisor

Evaluating Response to Tuberculosis Therapy: A Generalized Estimating Equation Model

by

Ryan Patrick Kyle
Bachelor of Science
University of Washington
2003

Faculty Thesis Advisor: David G. Kleinbaum, Ph.D.
Thesis Field Advisor: Charles M. Heilig, Ph.D.

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

2012

Abstract

Evaluating Response to Tuberculosis Therapy:
A Generalized Estimating Equation Model

by Ryan Patrick Kyle

A standard course of therapy for pulmonary tuberculosis (TB) requires six to nine months of medication. Shortening the time required for stable cure is a priority in current TB research. Current Phase 2 trials are large and expensive, given a shortage of surrogate markers for therapy response and small differences in treatment effect between current candidate drugs. Improved identification of those less responsive to therapy may reduce the time and cost needed to identify more effective treatments. The author examined data for 531 participants from TB Trials Consortium Study 29, a randomized, open-label, multicenter clinical trial comparing rifapentine versus rifampin administered 5 days per week during the first 8 weeks of combination therapy for pulmonary TB. Individuals without liquid culture results at baseline or at the end of intensive phase therapy were not included in the analysis; the final sample included 375 participants. The probability of negative culture in liquid media by week of therapy and treatment arm was obtained using a generalized estimating equation model that adjusted for geographic region, presence of productive cough at baseline or by week of therapy, cavitation at baseline, presence of diabetes, HIV status, cigarette use history, bilateral involvement at baseline, days to detection on MGIT 960 at baseline, sex and baseline smear grade. Presence of productive cough by week (adjusted odds ratio = 0.58; 95% confidence interval: 0.42, 0.81) and decreased days to detection at baseline (aOR = 1.09; 95% CI: 1.02, 1.16) were significantly associated with delayed culture conversion. Interactions between weeks on therapy and diabetes or history of cigarette use were identified. These findings may suggest that further clinical examination of the relationship between TB, diabetes and cigarette use is needed.

Evaluating Response to Tuberculosis Therapy: A Generalized Estimating Equation Model

by

Ryan Patrick Kyle
Bachelor of Science
University of Washington
2003

Faculty Thesis Advisor: David G. Kleinbaum, Ph.D.
Thesis Field Advisor: Charles M. Heilig, Ph.D.

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

2012

ACKNOWLEDGEMENTS

This research was supported by the Division of Tuberculosis Elimination at the Centers for Disease Control and Prevention, and a CDC Foundation fellowship made possible by a grant from Sanofi-Aventis.

❧

I would like to express my gratitude to my advisor, David Kleinbaum, whose excellent advice and thoughtful comments have done much to improve the quality of this thesis.

❧

I am indebted to Bill Mac Kenzie, whose kindness, humor, and reason have enriched my experiences considerably. One day I hope to be a "simple country epidemiologist" myself.

❧

Most of all, I offer thanks to Chad Heilig, my mentor and friend. Your generosity of spirit, moral support and strong sense of scientific rigor have had a profound impact on my graduate education. The example you set is a hard one to follow, but I will try.

DEDICATION

To my father and grandmother
To my family – your love and encouragement have made all things possible
To my fiancée Kara – I look forward to each day we share

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Background

1.1 Epidemiology of tuberculosis

Tuberculosis, a pulmonary infection caused by *Mycobacterium tuberculosis* (Mtb), remains a significant cause of mortality and morbidity globally. In 2007, 1.77 million deaths were attributed to Mtb, a figure surpassed only by deaths due to HIV infection (1). The disease is transmitted via exposure to inhaled droplets containing Mtb, with dispersal facilitated by talking, singing, coughing and sneezing (2). Of those who are exposed and infected, an estimated 5-10% will eventually develop active disease (3). Those who are asymptomatic, but harbor viable bacteria in their lung tissues, are considered to be latently infected; approximately one in three individuals worldwide is in this category (4). Domestically, the incidence of tuberculosis is declining. In 2010, the incidence of 3.6 cases per 100,000 population represented a 3.9% decrease from the previous year (5). While promising, the incidence among foreign-born residents highlights additional challenges. In 2010, the number of new TB cases among foreign-born residents was 11 times greater than for U.S.-born individuals (5). Globally, the distribution of Mtb is also unbalanced, as the majority of cases are found in 22 high-burden countries, largely in Asia and Africa. Nearly half of all cases annually are identified in China, Bangladesh, India, Indonesia and Pakistan (2). Assuming

the current annual rate of reduction in incidence remains unchanged from 1%, the goal of tuberculosis elimination is unlikely to be achieved prior to 2050 (6). Better strategies and new drug regimens are needed to achieve greater success in the prevention and treatment of pulmonary tuberculosis.

1.2 Challenges in the treatment of pulmonary tuberculosis: the problem of persistence

Over 50 years after the introduction of the antibiotic isoniazid, the successful treatment of pulmonary tuberculosis remains difficult. Compared to other respiratory infections, a course of combination therapy for disease due to Mtb is prolonged. Current guidelines recommend the use of a six to nine month course of drugs to achieve cure of pulmonary tuberculosis (7–9). The requirement for prolonged therapy to treat the disease is largely attributable to the behavior of Mtb within the human host. The resilient nature of Mtb results from its capability to transition from active growth into a state of vastly reduced growth and metabolic activity in times of environmental stress and evasion of immunologic processes. Little is known about the metabolic processes of dormant mycobacteria, although current evidence indicates that metabolism slows considerably and growth is intermittent (10). As a majority of antibiotic drug targets include metabolic or cell growth pathways, the elimination of such bacteria poses a significant problem. Even for those drugs considered highly effective at killing persisting mycobacteria, eradication of remaining bacilli often requires several months.

The results of several quantitative microbiologic studies suggest that Mtb populations may be dichotomized into actively replicating bacteria and nonreplicative persisters (11, 12). Recent reviews by Mitchison and Zhang propose that the latter class of bacteria may be further subdivided into multiple stages of persisters, although current data regarding this distinction is limited (10, 13, 14). The modern course of antibiotic treatment and

corresponding response to therapy both follow a biphasic pattern. Drugs administered up to two months after the start of treatment are said to be part of an intensive phase regimen, while drugs taken thereafter form the continuation phase of therapy (15).

Effectiveness of anti-tuberculosis drugs is generally characterized in two phases: early bactericidal activity (EBA) and sterilizing activity. EBA represents the potential of an antibiotic to eliminate actively dividing bacteria in sputum that grow on solid media during the initial 7–14 days of treatment (13, 16). Sterilizing activity describes the ability of a drug to kill persisting bacteria, by sputum sterilization of organisms that cannot be grown by routine culture. In contrast to EBA, the sterilizing activity of currently available TB drugs occurs over several months of treatment, and is effectively the rate-determining step in achieving stable cure of Mtb (17).

An examination of EBA requires repeated measurements of the population of bacteria present during weeks one and two of therapy. Estimation of bacillary load is typically performed by quantitatively culturing samples of patient sputum on solid media. Once successive negative cultures have been obtained for an individual receiving treatment for culture-confirmed Mtb, the patient is said to have experienced stable culture conversion. Shortening the course of Mtb therapy will require the use of drugs with more potent sterilizing activity than those now in use.

1.3 The pursuit of superior sterilizing activity

A current focus in TB drug development is on identifying new drugs capable of reducing the time to culture conversion, and thereby decreasing the length of therapy. These efforts are complicated by the rising costs of clinical trials. A recent analysis suggested that 30 compounds in Phase 1 testing would be required to identify one successful drug for Phase 2 evaluation, at a potential cost of $400 million and 12 years of development (18). An earlier survey of 10 pharmaceutical firms estimated an even higher cost for new drug discovery of $802 million, adjusted for inflation (19). Fewer than 10% of candidate compounds proceed

from Phase 1 clinical trials to final approval for production (20). The expensive nature of drug research and development is better understood after considering the history of tuberculosis clinical trials.

Early studies on the treatment of tuberculosis were conducted by the British Medical Research Council (BMRC), over a 40-year period from 1946 until 1986. The first published BMRC randomized controlled clinical trial was undertaken in 1946, and compared the effect of bed rest with that of streptomycin treatment plus bed rest in 107 tuberculosis patients (21). In 1952, the BMRC compared the effect of three regimens: isoniazid alone, streptomycin with 4-aminosalicylic acid (PAS), and streptomycin plus isoniazid. As in the 1946 study, the number of trial participants was small by modern standards, at 331 participants. Given that few effective treatments were available at the time of the early BMRC studies, small trials were sufficient to characterize large differences in the effect of candidate drugs (22). Over time, as the potency of treatments increased, larger sample sizes were required to demonstrate statistically significant superiority of drugs given greater than 90% efficacy of current standard Mtb therapy (23).

While production of anti-TB drugs with superior sterilizing activity is essential to the development of better regimens, there are further constraints. These compounds should be able to kill bacteria in all metabolic states rapidly. In addition, a new treatment must have an extended safety profile, given the long duration of therapy required (24). Demonstrating safety and efficacy is both time-consuming and costly. In lieu of dramatically increased funding for TB research, more efficient methods of evaluating new drug candidates are required.

1.4 Improving methods for the evaluation of new regimens: surrogate markers of sterilizing activity

Reducing the time and expense required to identify promising new regimens is complicated by the lack of validated surrogate markers of response to therapy. Surrogate markers have been defined by Katz as "laboratory measurements or physical signs used in therapeutic trials as a substitute for a clinically meaningful endpoint that is expected to predict the effect of therapy" (25).

Table 1.1: Surrogate markers of response to therapy

Biomarkers under evaluation	**Bactericidal activity**	**Sterilizing activity**	**Limit of detection**
Time to detection of growth in an automated broth culture system	✓	?	≤ 1 CFU/mL
Two month sputum culture conversion in liquid media	?	?	≤ 1 CFU/mL
Serial sputum colony counting	✓	?	10^1-10^3 CFU/mL
Quantitative nucleic acid amplification of Mtb in sputum or blood	?	?	100 CFU/mL
Two month sputum culture conversion on solid media	✓	✓	10^1-10^3 CFU/mL
Two month sputum smear	✓		4 AFB/100 HPF
Unevaluated potential biomarkers			
Volatile organic compounds			
Macrophage activation markers			
Multiple host markers Proteomics, transcriptomics			

References: (26–29). Abbreviations: CFU = colony forming unit, IU = international unit, AFB = acid-fast bacilli, HPF = high power field.

1.4.1 Two-month culture conversion

While culture status on solid media following two months of treatment remains the only evidence-based surrogate for stable cure, the use of this outcome for modern drug trials is problematic (30). Given that current culture conversion rates for standard TB therapy are high (> 80%), a large sample size is necessary. Studies examining rates of two-month culture conversion on solid media require 100 to 150 patients per study arm, resulting in longer, more expensive trials (31). In addition, two-month culture conversion produces no data on the effect of treatment prior to reaching the endpoint (32). This important detail is a primary obstacle in efforts to shorten the course of Mtb therapy. The current approach is also an inefficient use of available information, as only the presence or absence of mycobacteria is assessed. Counts are effectively dichotomized into a binary value, resulting in a potential loss of statistical power (17). Because of each assay's limit of detection, a sample may appear negative, although viable bacilli remain at a level potentially undetectable by the assay. This latter detail has important implications for individuals who later experience treatment failure or relapse.

A more ideal surrogate for bacillary load should be measured on a continuous scale. Such a measure would permit more frequent comparisons of sterilizing activity between differing drug regimens or antibiotics during treatment. Relative to a binary outcome, a continuous measure of sterilizing activity would allow for the examination of trends and variability between individuals during therapy. The limitations of two-month culture conversion have resulted in several biomarkers being proposed as clinical endpoints. Several surrogate markers of response to therapy have been examined clinically, and additional biomarkers have been proposed but not yet evaluated in a clinical trials setting (table 1.1). Time to stable conversion during therapy is expected to serve as a superior surrogate for sterilizing activity, and has been examined in studies conducted by the Tuberculosis Trials Consortium (TBTC). However, the results of several clinical trials suggest that time to stable conversion may be highly variable between several participant groups (33).

1.4.2 Sputum smear

Sputum smear grading under microscope is a commonly used biomarker of disease status and response to therapy, and may be the only viable option in resource-limited settings. For patients with productive cough, sputum specimens are collected, stained, and viewed using conventional or fluorescence microscopy. Smear grading is then used to categorize the number of acid-fast bacilli per high power field. Domestic smears are graded from 1+ to 4+ using the CDC scale, while internationally smear grades 1+ to 3+ are assigned using the International Union Against Tuberculosis and Lung Disease grading scale (34). As the quantity of bacteria within the lungs increases, the probability of sampling a larger number of bacteria in sputum specimens also increases.

Sputum smear grading is frequently used as a semi-quantitative measure of bacillary load. Clinical results indicate that the risks of recurrence, mortality and morbidity are typically higher for individuals with higher grade smears at the start of therapy compared to those with lower smear grades (35). This relationship is potentially confounded, however, among the elderly and those with HIV infection. For these individuals and others with disseminated disease not confined to lung tissue, sputum smear grading may underestimate the burden of infection. Increased time to culture conversion has been associated with medium and high smear grades in previous studies (35, 36). In a systematic review of 28 studies, Horne et al. argue that sputum smear is an inadequate marker of poor treatment outcomes (37).

1.4.3 Serial sputum colony counting on solid media

In contrast to smear grading, quantitative sputum culture on solid media permits an estimation of both bacillary load and response to therapy. Serial sputum colony counting (SSCC) employs serial dilutions of sputum, plated on solid media selective for growth of Mtb, for quantification of colony forming units (CFUs) per milliliter of sputum. Although SSCC was originally developed as a measure of EBA, it is now considered a promising candidate for

assessment of sterilizing activity. Rustomjee et al. successfully demonstrated the application of SSCC in an open-label, 8-week Phase 2B clinical trial of ofloxacin, gatifloxacin, and moxifloxacin (17).

An attractive feature of SSCC is the reduced study size required to identify small differences in therapeutic effects. Compared to the large sample sizes per arm required in trials using 2 month culture conversion as an endpoint, SSCC may require as few as 50 participants per arm (38). There is also the potential for more efficient factorial designs given SSCC as an endpoint, which would permit the examination of dose escalation and companion therapies in a variety of combinations (38). In addition, the results of nonlinear mixed effects analysis of SSCC data from several studies suggests that this method yields improved precision compared to other approaches (17, 39).

However, there are limitations in the use of SSCC repeated measures analysis. Supervised sputum collection is necessary, making intensive participant follow-up essential (40). The optimized sampling scheme required is complex, with up to 10 time points required to realize the potential gain in power the method permits (38). The preparation of samples is also important, as some have questioned the effects of non-selective decontaminants (such as NaOH) on sample viability (39, 40).

1.4.4 Automated liquid culture systems

The advent of liquid culture systems for the detection of mycobacteria in sputum specimens has enhanced and standardized several aspects of data collection in TB clinical trials, particularly in high-burden countries. Liquid culture provides high sensitivity and reasonably fast detection of growth, and an affordable means of differentiating between various species of mycobacteria. In anticipation of wider use of the technology, the World Health Organization published revised policies in 2007 regarding the use of automated liquid culture systems.

The BD (Becton, Dickinson and Company) BACTEC Mycobacterial Growth Indicator Tube 960 (MGIT) system analyzes oxygen metabolism by bacteria by detection of decreased

oxygen concentration via fluorescence. The fully automated system measures fluorescence within sample tubes at hourly intervals. Once the oxygen concentration within an indicator tube decreases beyond a specified threshold, growth is presumed and the instrument records the time to detection. If no fluorescence is detected within 42 days, testing concludes and no growth is recorded.

The MGIT 960 system offers multiple advantages relative to solid culture. The continuous nature of time to detection suggests a convenient relationship with the actual size of the Mtb population within a sputum sample. As the quantity of active bacteria increases, the rate of oxygen consumption increases and the time to detection decreases; the inverse is also presumed to be true. While SSCC may offer similar information, the technique is more complicated. The results of studies which obtained both liquid and solid cultures indicate that culture-positivity persists longer in liquid media than on solid media (17). As a result, liquid culture may be more appropriate for examining persister populations than solid culture. Some questions have been raised regarding the restricted nature of liquid culture relative to SSCC as a potential endpoint in clinical trials, the differences in results between various detection systems, and the role of time-dependent variations in bacterial growth in interpretation of the results (32).

1.4.5 Nucleic acid amplification

In contrast to culture-based methods, nucleic acid amplification (NAA) detection systems permit more efficient diagnosis of Mtb infection and drug sensitivity. While a culture-based method may take several weeks before indicating specimen positivity, current NAA instruments can produce results in less than two hours. The Cepheid Xpert Mtb/RIF assay has demonstrated accuracy as a diagnostic in clinical settings (41). This particular NAA assay amplifies the rifampin resistance-determining region of the *rpoB* gene, which codes the beta subunit of RNA polymerase within *M. tuberculosis*, using a real-time polymerase chain reaction. In experiments performed by Blakemore et al., the assay was highly sensitive,

correctly identifying 79 of 79 Mtb sample strains as positive, and also highly specific as no false positives were detected among 89 samples (42). The limit of detection for the Cepheid Xpert Mtb/RIF has been estimated at 100 to 131 colony forming units (CFU) per milliliter of sputum, an order of magnitude higher than that of current automated culture-based systems, but better than smear culture techniques (29, 43).

While NAA holds considerable promise for efficient and inexpensive diagnosis of Mtb infection and rifampin resistance in clinics and hospitals, its potential to measure sterilizing activity remains unknown. Unlike broth culture systems, it currently provides limited quantitative information to the user which could be used as a surrogate marker of bacillary load. Further, because Xpert Mtb/RIF uses a real-time PCR-based method of identifying nucleic acid sequences, any quantitative results produced by the assay must control for the presence of dead bacteria. In contrast, culture-based systems only measure viable bacteria. In a comparison of Xpert Mtb/RIF with other nucleic acid technologies, Scott et al. suggest that although the system correctly identified 96% of smear-positive/culture-positive specimens, the success rate decreased to 61% for smear-negative/culture-positive specimens (44). This may suggest that the current version of the Xpert Mtb/RIF assay is less effective at identifying samples with smaller quantities of bacteria. For the moment, NAA may be better suited as a tool to ascertain infection and detect antibiotic resistance in new patients, rather than a means to repeatedly quantify bacillary load in a experimental trials setting.

1.4.6 Biomarkers yet to be used in clinical trials

Gas chromatography and mass spectrometry (GC-MS) has also been proposed as a rapid diagnostic method for presence of Mtb (45, 46). A review of the literature indicates that GC-MS has yet to be used in a clinical trial. GC-MS may enable much faster detection of Mtb than is possible with solid or liquid culture. Analysis of Mtb specimens with GC-MS has revealed gases and volatile compounds essential in biofilm formation, thought to be a key environmental requirement for persistence (47). Given further refinement, GC-MS

may identify organic compounds produced by both persister and non-persister classes of mycobacteria. However, the current lack of published, peer-reviewed methods to reliably estimate bacillary load from a GC-MS analysis limits the current utility of the technique.

A more recent development is the molecular bacterial load (MBL) assay, proposed as a culture-free biomarker of Mtb bacillary load during therapy by Honeyborne et al. The authors suggest that this method permits the detection of live bacilli in sputum through an assay of Mtb 16S rRNA, producing results more rapidly than culture-based methods, and more simply than is possible by SSCC (48). Since MBL analyzes the presence of short-lived RNA species, it may allow the quantification of only viable bacilli without the use of culture. While additional analysis and application of this new tool is necessary, it may be of considerable utility in future clinical trials.

1.5 Factors affecting response to therapy

1.5.1 Cavitary tuberculosis

The formation of cavities among those with pulmonary tuberculosis further complicates treatment and prevention. Cavities are gas-filled spaces produced by the destruction and expulsion of dead lung tissue within pulmonary lesions, and may be seen in chest radiography (49). Mtb is generally free to proliferate in cavities, relatively isolated from the human immune system. Cavities may also affect transmission, as they facilitate the growth of large numbers of organisms and aerosolization of bacteria, increasing the risk of transmission (49). Radiographic evidence of cavitation is suggestive of advanced disease and a potential predictor of delayed time to stable conversion (50, 51). Cavitary tuberculosis has also been identified as a potential predictor of treatment failure and relapse (52).

1.5.2 Diabetes

Diabetes is also believed to affect response to therapy and time to culture conversion. Host immune response to Mtb among diabetics appears to be impaired relative to non-diabetics (53). Recent research suggests that time to stable conversion among diabetics is longer than for non-diabetics (33, 54, 55). A retrospective cohort study by Dooley et al. proposed that the odds of mortality among patients with diabetes may be twice that of patients without diabetes (55). In addition, multiple longitudinal cohort and case-control studies determined that the odds of developing Mtb among diabetics are two to eight times that of non-diabetics (53).

1.5.3 Smoking

Although the biological mechanism by which smoking increases susceptibility to tuberculosis continues to be debated, research by Kumar and colleagues found that simultaneous sensing of carbon monoxide, nitric oxide and oxygen is an essential component of dormancy regulation in Mtb (56). Carbon monoxide is a well-known byproduct of tobacco smoking, and levels of the gas in the lungs and breath may be measured to identify recent smokers (57). The risk of Mtb among smokers has been estimated at two to three times that of non-smokers (58). It remains unclear whether the elevated risk of infection among tobacco smokers results from generation of larger populations of persisting bacteria after initial infection, through immune suppression or damage within pulmonary tissues, or a synergy of these factors (59, 60). Slama et al. reported that strong evidence existed for a dose-response effect of smoking on the risk of Mtb infection in their systematic review and meta-analysis of 53 studies (61). The results of a study by Güler et al. identified significant associations between history of prior smoking, diabetes, age and sex with delayed culture conversion (54).

1.5.4 Geographic region

A relationship between geographic region and time to culture conversion has been observed in clinical trials. The results of TBTC studies 27 and 28 indicate that time to culture conversion in liquid media for patients at African sites was delayed relative to that of patients at non-African sites (33). The primary analysis of TBTC Study 28 data suggested that diabetes and HIV status, smoking, age, race and baseline severity of disease did not explain the extended time to culture conversion amongst participants in African countries.

1.6 Summary

Although considerable progress has been made in describing the biology of *M. tuberculosis*, a reliable method for the estimation of bacillary load during a full course of therapy is not currently available. Better surrogate markers of response to therapy are needed, both to shorten the course of treatment and of future clinical trials. More comprehensive statistical models incorporating several biomarkers, both culture-based and culture-free, as well as non-microbiological predictors of response to therapy, could be very useful to clinicians, epidemiologists and laboratorians.

Chapter 2

Manuscript

2.1 Introduction

Although access to antituberculosis medications is gradually improving, effective management of infection due to *Mycobacterium tuberculosis* (Mtb) remains a major challenge. Stable cure, defined as the attainment of a negative culture followed by no subsequent positive cultures, generally requires a prolonged treatment period. Typically, Mtb treatment regimens require at least three medications taken simultaneously for six to nine months. Culture conversion, an intermediate step to stable cure and a common marker of response to therapy, is defined as attainment of a negative mycobacterial culture following a previous positive culture. Ensuring complete compliance with Mtb treatment is difficult, particularly in areas with limited health services. While implementation of the World Health Organization DOTS (directly observed treatment, short-course) strategy has improved treatment success rates, treatment failure, relapse and drug resistance remain primary concerns in the effort to stop the spread of Mtb.

Recent advances in drug research have yielded novel treatment options; the efficacy of these new medications remains to be seen. While promising, widespread adoption of new anti-Mtb regimens is unlikely to occur in the immediate future. Further, Phase 2 clinical

trials of novel candidates for Mtb therapy are quite expensive. More rapid identification of effective treatments is essential for improving future tuberculosis trials. A deeper, more quantitative characterization of the dynamics of Mtb infection associated with treatment failure and relapse, could provide a useful tool to clinicians, epidemiologists and those engaged in laboratory research.

To better characterize sterilizing activity of Mtb drugs, more accurate methods of estimating bacillary load in patients with pulmonary tuberculosis are required. The improved identification of individuals less responsive to treatment could help decrease current trial sizes, as it would enable better discrimination of treatment effects between study participants. In efforts to support these objectives, I propose a statistical model to consider the effect of several predictors on mycobacterial liquid culture status and response to therapy. The primary aim of this research is to determine and describe any difference in treatment outcome between the two treatment arms in the study data I have analyzed, as indicated by culture results during intensive phase treatment. Secondary research goals include identification of potential confounders and effect modifiers of response to therapy.

2.2 Methods

2.2.1 Study population

Data used in this analysis were obtained in Tuberculosis Trials Consortium Study 29 (http://www.clinicaltrials.gov/ registry number NCT00694629), a randomized controlled trial comparing the antimicrobial activity and safety of rifapentine versus rifampin administered 5 days per week during the first 8 weeks of combination therapy for pulmonary tuberculosis. Study 29 was a prospective, multicenter, open-label study. Participants were enrolled at 24 sites in multiple regions (16 sites in North America, 3 in South Africa, and individual sites in Brazil, Peru, Spain, Uganda and Vietnam). A total of 531 participants were enrolled, with 529 receiving any study therapy. The primary endpoint was negative

sputum culture on liquid or solid media at completion of intensive phase therapy. To be eligible for inclusion in the study, participants were required to be adults with suspected pulmonary Mtb and acid-fast bacilli (AFB) in a sputum specimen. Participants meeting any of the following criteria were excluded: pregnancy, currently breastfeeding, body weight less than 40 kilograms, central nervous system TB, pulmonary silicosis, allergy or intolerance to any study drug, current or planned therapy during study treatment with antiretroviral medications (in addition to cyclosporine or tacrolimus) and initial sputum cultures negative for Mtb or with growth of an Mtb strain resistant to rifampin, isoniazid, or pyrazinamide. All study participants were tested for HIV.

Study participants were randomly assigned to receive rifapentine or rifampin (both provided at approximately 10 mg/kg), administered once daily for 5 days per week. In addition, all participants also received isoniazid, pyrazinamide, ethambutol and pyridoxine during intensive phase therapy, the first 8 weeks of treatment. Due to the absence of an available rifapentine placebo at the time of this trial, rifapentine and rifampin administration was not blinded. Stratified randomization was used, according to presence of cavitation at baseline and region of enrollment (North America, South Africa, South America, Spain, and Uganda). Participants receiving 40 directly observed therapy (DOT) doses in 54 to 70 days were identified as having completed study therapy. Blood and sputum specimens were collected at baseline and every two weeks thereafter until completion of intensive phase therapy; at 8 weeks, two sputa were obtained from each participant. Liver function was assessed by evaluation of serum alanine transaminase, bilirubin, and creatinine levels. Cultures were performed and processed at local site laboratories on both Lowenstein-Jensen (LJ) solid media and BACTEC Mycobacterial Growth Indicator Tube (MGIT, Becton Deckinson and Co., Franklin Lakes, NJ) liquid media using the MGIT 960 system.

Individuals with a baseline specimen negative for growth of mycobacteria, or whose specimen was found to be resistant to isoniazid, rifampin, or pyrazinamide were excluded from classification into a modified intention-to-treat (MITT) group. A further subset of

the MITT group included only protocol-correct participants. To be protocol-correct, a participant was required to have completed study therapy within 54 to 70 days and with an evaluable culture at the end of intensive phase treatment. Analysis of both protocol-correct and MITT groups was specified in the study protocol. Following consultation with the institutional review board at Emory University, it was determined that no IRB review was required for the analysis presented in this thesis.

2.2.2 Data analysis

Generalized estimating equation (GEE) models were used in the analysis, as this method permits estimation of population-level averages, using a repeated measures approach that also considers correlations within clusters of observations (62). The GENMOD procedure in the SAS software package was used for model fitting (v 9.2, SAS Institute, Cary, NC). Additional analysis was also performed using R (v 2.14, R Development Core Team, Vienna, Austria). The specified outcome for all models was a negative culture result on liquid media, as identified by no detectable Mtb growth after 42 days using the MGIT 960 system.

Liquid culture status was examined during weeks 2, 4, 6 and 8 of intensive phase therapy, with treatment arm and weeks of study treatment considered as exposure variables. Relationships between potential confounders, the study outcome and either exposure were first examined in multivariable GEE models including the two exposure variables and partially adjusted for baseline cavitary status and geographic region, which were used to stratify treatment allocation. Baseline cavitary status and geographic region were also retained in all subsequent models given their nature as study design variables. GEE model selection proceeded according to a multistage process: variable screening, interaction assessment, analysis of confounding, and comparison of relative precision were used to identify a preferred model.

A total of 11 variables of potential interest were identified prior to screening, not including the two study design variables (table 2.1). Correlation information criterion (CIC) values were used to select a correlation structure using a complete preliminary model containing

all variables, while considering the biological plausibility of the various choices (63, 64). Five working correlation structures were considered: exchangeable, independence, Toeplitz (stationary 1-dependent), first order autoregressive, and unstructured. Collinearity was examined for this preliminary model using each of the five structures, in case model stability was dependent upon a chosen correlation structure. Collinearity diagnostics were performed by evaluating condition indices and variance decomposition proportions (VDP) obtained from the inverse of the Fisher information matrix with cutpoints of 30 for the condition indices and 0.5 for the VDPs (65, 66). Presence of a condition index greater than 30, with two non-intercept VDP values higher than 0.5, was determined to indicate a potential collinearity problem. Collinearity was remedied by eliminating one or more covariates with high VDPs while preserving model hierarchy. To be retained in the model, potential effect modifiers were required to have a P-value less than 0.05.

Given the large number of initial model terms, a screening process was used to eliminate three (age at study entry, HIV status, and history of prior tuberculosis) from consideration. A change-in-estimate approach assessed the effect of excluding these terms from the preliminary model; as their omission did not result in a meaningful change in effect estimates, they were excluded from further consideration.

In the next stage, effect measure modification was assessed by modeling all potential product terms including either exposure and all other variables using a backward elimination approach; only pairwise interactions were examined. If a product term was found both to be significant and of interest, it was retained in all later models. Lower order constituents of significant product terms were also retained to preserve model hierarchy.

After interaction assessment, an analysis of confounding was performed by modeling all possible subsets of variables retained as potential confounders (67). The model identified after assessment of interaction was used as the initial model in the analysis of confounding. In case of significant interactions, a table of odds ratios was consulted to evaluate effect estimates at multiple levels of the observed interaction. If an estimated odds ratio in the

resulting table differed by more than ±10% from the initial model's corresponding result, that model was determined not to control adequately for confounding. Following this process, any models which were determined to adequately control for confounding were eligible as candidates for precision assessment and comparison with the initial model.

In the final stage of model selection, models which adequately controlled for confounding were ordered according to relative precision as evaluated by comparing a given model estimate's standard error against that of the initial model. If a candidate model produced a meaningful gain in precision, that model was chosen as the preferred model. Otherwise, the inital, fully adjusted model was chosen instead.

A type I error rate of 0.05 was used for all statistical tests of significance. For analysis of categorical data, a Pearson's chi-squared test for homogeneity of proportions was performed. Student's t-test was used to compare mean values between groups for normally distributed variables with comparable variances. For comparisons where variances were unequal or data were not assumed to be normally distributed, the Wilcoxon rank sum test was used as a nonparametric alternative. Generalized score tests were used to test the significance of model terms (68, 69).

2.3 Results

At the conclusion of enrollment, the study sample included 531 participants. Following the exclusion of 142 participants who were not determined to be protocol correct, an additional 14 participants were excluded from the analysis due to missing liquid culture results at baseline (figure 2.1). The final analysis examined data for 375 participants. Means and proportions for covariates included in the final model are summarized in stratified form according to treatment arm and geographic region (table 2.2).

Median age among participants at African sites was 28 (SD=9.2) years, compared to 42 (SD=15.3) years at non-African sites. This difference in median age by geographic region was

statistically significant (Wilcoxon rank sum test; P < 0.001). The median time to detection at baseline for African sites was 6.2 (SD=2.9) days, versus 8.3 (SD=4.0) days for non-African sites; this difference was also statistically significant (Wilcoxon rank sum test; P < 0.001). The distribution of days to detection at weeks two and four among participants was clearly right-skewed. However, by weeks six and eight of therapy, days to detection among participants was more symmetric (figure 2.2).

Presence of productive cough at baseline was reported by 89.1% of participants, and common throughout all four strata. Diabetes was infrequently reported, particularly by subjects at African sites, where only one individual indicated a history of diabetes at baseline. Overall, 40.8% of participants reported a history of any prior cigarette use. The prevalence of previous cigarette use was significantly lower among subjects at African sites than at non-African sites, with 58.2% of all smokers in the latter group (Pearson's chi-squared test; P < 0.001). At the end of intensive phase therapy, 106 of 375 participants remained positive for Mtb on liquid culture.

2.3.1 Model selection, and assessment of interaction and confounding

An analysis of collinearity was performed for the preliminary model using each of the five proposed correlation structures. No potential collinearity problems were detected following these diagnostics. A subsequent comparison of CIC values suggested that either an unstructured or a first order autoregressive correlation structure would be reasonable given the data. The correlation structures with the smallest CIC values were considered more appropriate. A first order autoregressive correlation structure was chosen for all fitted models, both due to the longitudinal nature of the study and for greater parsimony. A detailed explanation of the CIC and methods for its computation are available in Appendix A.

The results of an exploratory multivariable analysis suggested that days to detection on the MGIT 960 system, sex, baseline smear grade, and presence of productive cough at baseline and week of therapy were all significantly associated with culture status (table 2.3).

Bilateral involvement on baseline radiograph, diabetes or HIV status, age, and history of any cigarette use were not found to be significantly associated with culture status in partially adjusted multivariable models. History of cigarette use and diabetes status at baseline were deemed potentially important factors associated with the outcome, following a review of recent clinical research which found these to be effect modifiers of response to therapy (33).

As the all-possible-subsets approach requires specifying 2^n combinations of models when n variables are considered, a screening process was used to further reduce the problem space. A change-in-estimate approach subsequently eliminated HIV status, age at study entry, and history of prior tuberculosis from consideration. The resulting model included both exposures, the study design variables, and six other terms: bilateral involvement at baseline, baseline days to detection on MGIT 960, productive cough at baseline and at weeks 2 through 8 of therapy, baseline smear grade, and sex. Smear grade at baseline was treated as a categorical rather than ordinal variable; smear grades under microscope were classified as low, medium and high according to the number of bacilli present per high power field. Specimens with 1-9 bacilli per 10 high power fields were graded low, those with 1-9 bacilli per high power field were graded medium, and specimens with more than 9 bacilli per high power field were classified as high.

After the screening procedure, an assessment of effect measure modification was performed. A second analysis of collinearity was conducted, given the large number of product terms present prior to the start of backward elimination. No collinearity issues were identified using any of the five previously proposed correlation structures in this second round of diagnostics. Following selection by backward elimination, two significant interactions of interest were identified between weeks on therapy and diabetes or history of cigarette use. The two product terms associated with these interactions and their constituent variables were retained in all later models.

Since the model included two product terms, a table of odds ratio estimates was constructed for each of the 63 models under consideration. Given that weeks on therapy was

derived from the number of days since the start of therapy, it was analyzed as a continuous variable. As a result, comparisons were made using a series of 2×6 tables with odds ratio estimates for the two interactions using three levels of weeks on therapy (3, 5, 7 weeks, as described by tertile). Three models controlled adequately for confounding relative to the full, initial model, but none provided meaningfully better precision, assessed by examination of a table of confidence intervals and comparison of standard errors for exposure terms. As a result, the chosen model adjusted for all six covariates, the two product terms and their constituent variables, as well as baseline cavitary status, and geographic region (table 2.4).

2.3.2 Selected GEE model

No significant difference in the odds of culture conversion between the two arms was detected (adjusted odds ratio 1.32; 95% CI 0.91, 1.91; P = 0.15). Among participants reporting diabetes at baseline, the odds of culture conversion were lower than those not reporting diabetes in weeks two (aOR 0.43; 95% CI 0.16, 1.15; P = 0.09) and four (aOR 0.81; 95% CI 0.41, 1.58; P = 0.53) of therapy. Conversely, the odds among diabetics in weeks six (aOR 1.50; 95% CI 0.72, 3.13; P = 0.28) and eight (aOR 2.80; 95% CI 0.93, 8.43; P = 0.07) were greater than those not reporting diabetes at study entry. The odds of culture conversion among those indicating any previous cigarette use in weeks two (aOR 1.92; 95% CI 0.96, 3.87; P = 0.07) and four (aOR 1.20; 95% CI 0.74, 1.96; P = 0.46) of therapy were higher than those who did not report any previous cigarette use. In weeks six (aOR 0.75; 95% CI 0.50, 1.13; P = 0.17) and eight (aOR 0.47; 95% CI 0.28, 0.80; P = 0.005) of intensive phase therapy, culture conversion among those who had ever smoked appeared to be lower than for those who had never smoked before.

Participants at African sites had a much lower probability of culture conversion compared to those at non-African sites (aOR 0.40; 95% CI 0.25, 0.62; P < 0.001). In addition, the odds of a negative culture were 72% lower for participants reporting productive cough at baseline compared to participants reporting no productive cough (aOR 0.28; 95% CI

0.14, 0.57; P < 0.001). The odds of culture conversion for individuals with productive cough at a given week of therapy were significantly lower than for those not indicating productive cough (aOR 0.60; 95% CI 0.43, 0.83; P = 0.003). The odds of positive culture for individuals with cavitary disease were 34% lower than for individuals with no cavitation at baseline (aOR 0.68; 95% CI 0.46, 1.02; P = 0.07).

As suggested by the partially adjusted multivariable analyses, days to detection, presence of productive cough, and smear grade were all significant baseline predictors of culture conversion status. Given the model, the odds of a negative culture increased by 9% for each additional day increase until detection on liquid culture of sputum at baseline (aOR 1.09; 95% CI 1.02, 1.16; P = 0.02). A trend in decreasing odds of culture conversion was observed with respect to increasing baseline smear grade. Failure to attain a negative culture at each time point during intensive phase therapy was associated with higher initial bacillary load as indicated by medium and high baseline smear grades. Sex was also found to be a significant predictor of culture status. The odds of a negative culture on liquid media for female participants were 66% higher than those of male participants (aOR 1.66; 95% CI 1.11,2.50; P = 0.02). Bilateral lung involvement was not found to be significantly associated with culture status.

2.4 Discussion

Previous analyses of data collected in Study 29 employed logistic regression and survival analysis methods. This thesis is the first analysis of Study 29 data using a repeated measures approach. While the analytic methods differ, several findings presented here mirror those of the primary analysis of Study 29 data, which also examined demographic and biological predictors of culture status at end of intensive phase therapy.

As in this analysis, high bacillary load on baseline sputum culture and previous cigarette use were found to be significantly associated with culture status, and potential risk factors for delayed culture conversion. Conversely, female sex was associated with a decreased time

to culture conversion in both analyses. Furthermore, the primary analysis of Study 29 data did not find productive cough at baseline to be significantly associated with culture status at end of intensive phase therapy (aOR 0.78; 95% CI 0.33, 1.86; P = 0.58), but did suggest a significant association with age in years at study entry (aOR 0.98; 95% CI 0.96, 1.00; P = 0.02), which was not found to be significant in this analysis (70). While the interactions between weeks on therapy and either diabetes or history of cigarette use are potentially interesting, these findings must be considered carefully in the context of the data.

The measurement of cigarette use in this study did not consider multiple categories of smokers, including estimates of participant pack year history and whether or not participant was a current smoker or had ever smoked. Subsequently, it is unknown if individuals reporting any cigarette use at time of therapy were currently smokers, or how frequently they smoked. A more careful examination of the effect of cigarette use on time to culture conversion should consider the frequency of current smoking during intensive phase therapy. Given the results of previous research by Kumar et al., it is interesting to consider that smoking could potentially affect the time to culture conversion by promotion of dormancy in Mtb via exposure to exogenous carbon monoxide (56). It is reasonable to consider that the proportion of bacilli which are not able to be cultured with current diagnostics could increase as a result of carbon monoxide exposure, although examination of such an effect during therapy would probably be quite challenging.

The presence of diabetes was ascertained following review of concomitant conditions described on the participant history forms. This raises several important issues. The lack of fasting blood sugar or hemoglobin A1C data at enrollment and by week of therapy prevents an objective classification of diabetes, and an examination of a potential relationship between management of this condition and response to therapy. Further, among the 42 participants who reported having diabetes, only 9 also had data indicating use of insulin or other anti-diabetic drugs (metformin, glibenclamide, glucagon) at time of enrollment. No additional data regarding diabetes were available. While this may be attributable to under-

reporting of diabetes and associated medications, it is also a potential source of information bias. As a result, the measurement of diabetes in Study 29, and the significant interaction identified in this analysis, are both subject to a large degree of uncertainty.

For both diabetes and cigarette use, the distributions of the covariates differ measurably between African and non-African sites. As previous research suggests a significant difference in rates of culture conversion by region, this may limit the generalizability of these findings.

Compared to other selected predictors of culture status, the point estimates for the effect of female sex and history of previous cigarette use are relatively imprecise. A potential limitation of this GEE analysis is the inability to characterize subject-specific effects and between-participant heterogeneity; a random effects model would enable further examination of these factors among female participants and those reporting previous cigarette use.

The inclusion of symptoms at baseline and by week of therapy is a useful feature of this repeated measures analysis. Consideration of symptoms during treatment may incorporate additional, important information about response to therapy not captured by culture methods. Although the role of symptoms as time-varying and baseline predictors of culture status during Mtb therapy is not well understood, the results of this and an earlier analysis support further evaluation of these covariates. In addition to capturing additional information about response to therapy, the inclusion of productive cough either at baseline or as a repeated measure may convey details regarding sputum sampling. It is plausible that individuals with productive cough are more likely to produce larger volumes of sputa, which could potentially improve collection of viable bacilli within the lungs.

This thesis represents an intermediate step in a larger goal of the TBTC's efforts to better characterize bacillary load in those infected with pulmonary tuberculosis. Future efforts will examine days to detection on liquid media as a continuous outcome, using non-linear mixed models, as this is anticipated to provide a more precise estimate of individual bacillary burden.

Appendix A: Notes on GEE correlation structure selection

The correlation information criterion (CIC)

Model selection and comparison of correlation structures in GEE cannot be performed using criteria or heuristics developed from maximum likelihood principles, since GEE is not a likelihood-based method. As a result, methods such as Akaike's AIC and likelihood ratio tests should not be used. Pan proposed a generalization of AIC he termed the quasilikelihood under the independence model criterion (QIC) which can be used instead to compare and select a preferred correlation structure given a series of otherwise identical models:

$$QIC \equiv -2Q(\hat{\beta}(R); I, \mathcal{D}) + 2\,\text{trace}\,(\hat{\Omega}_I \hat{V}_r) \tag{1}$$

where Q denotes the quasilikelihood, $\hat{\beta}$ is a vector of regression coefficients to be estimated, $\mathcal{D}$ gives data used in the model, trace describes the sum of the matrix diagonal, $\hat{\Omega}$ refers to the inverse of the variance-covariance matrix A_I, and $\hat{V}_r$ represents the robust ("sandwich") covariance estimator. The QIC statistic is now available in most software packages supporting GEE. A review of recent research suggested that alternative methods for identifying appropriate correlation structures may provide superior performance. These alternatives include the correlation information criterion (CIC) proposed by Hin and Wang, the SC criterion described by Shults and Chaganty, as well as the Rotnitzky-Jewell criterion (63, 71, 72). Unlike the latter two criteria, the CIC is derived from Pan's QIC, including only the second term (while omitting the multiplier):

$$CIC \equiv \text{trace}\,(\hat{\Omega}_I \hat{V}_r) \tag{2}$$

Hin and Wang decompose the QIC and propose that the first term may be ignored when comparing correlation structures, as it does not depend upon the true intracluster correlation

structure (63). The authors suggest that the first term must be estimated with a degree of random error, and including it may mask the influence of the second term. The results of several simulation studies indicate that the CIC provides improved sensitivity in identifying the correct correlation structure, although not without limitations (63, 73, 74).

Given that the CIC does not penalize for overparameterization, there are potential issues associated with comparisons involving correlation structures with large differences in the number of correlation parameters being estimated (63, 75). The analyst should take into account the number of observations per cluster when using the CIC in its current form, and compare the number of correlation parameters in candidate structures. The potential for large differences is readily apparent when comparing two commonly used correlation structures in longitudinal analysis: the first order autoregressive structure (AR1), in which only one parameter is estimated, and the unstructured correlation structure, where $n_i(n_i - 1)/2$ gives the number of correlation parameters which need to be estimated. For the analysis presented in this thesis, an unstructured correlation structure requires only 6 parameters to be estimated, and the CIC remains a suitable method of comparing various structures. As an example, a study with three times as many observations per cluster (a total of 12 time points) would require 66 parameters to be estimated for a model using an unstructured correlation structure; in such a case, comparisons with other structures using the CIC may be potentially unreliable.

Computing the CIC in SAS

As of this writing, version 9.3 of SAS does not provide the CIC in output from the GENMOD procedure. Prior to version 9.2, the QIC and QICu were not available by default in SAS, and the qic.sas macro could be used to compute these statistics instead. I modified this program, which remains available online, to obtain the CIC statistic (76).

The code I modified, between lines 279 and 291, is provided here:

```
cic=trace(inv(oir)*vr);

/* Create output data set, OUTQIC= */
qic=qicr // qicu // cic;
criterion={"QIC" "QICu" "CIC"};
cname={"Value"};
create &outqic from qic [rowname=criterion colname=cname];
append from qic [rowname=criterion];
quit;

%if &syserr ne 0 %then %do;
 %put QIC: Error in computing QIC, QICu or CIC. Aborting.;
  %goto exit;
%end;
```

Computing the CIC in R

Jack Weiss has provided code, included below with his permission, which computes the QIC and CIC for models fit with `geepack`. The first argument is a model fit with any non-independence correlation structure, and the second argument is a model fit with the same covariates but an independence correlation structure.

```
QIC.binom.gee <- function(model.R, model.independence)
{
# calculates binomial QAIC of Pan (2001)
# author: Jack Weiss

AIinverse <- solve(model.independence$naive.variance)
V.msR <- model.R$robust.variance
trace.term <- sum(diag(AIinverse%*%V.msR))

# estimated mean and observed values and scale for binary data
mu.R <- model.R$fitted.values
y <- model.R$y
scale <- 1

# quasilikelihood for binomial model
quasi.R <- sum(y*log(mu.R/(1-mu.R))+log(1-mu.R))/scale
QIC <- (-2)*quasi.R + 2*trace.term
output <- c(QIC,trace.term)
names(output) <- c('QIC','CIC')
output
}
```

Appendix B: Code used to fit final GEE model

SAS code for final GEE model

The proposed model involves a binary outcome with no random effects. As a result, the GENMOD procedure was used to fit the model. A within-subject effect for visit number (1, 2, 3, 4) was specified to account for missing observations not occurring after the final recorded visit. If no within-subject effect for visit is provided, SAS assumes participant dropout, with all missing values occurring at the end of a cluster.

```
PROC GENMOD DATA=thesis.s29_gee_final
       CLASS arm (REF='rif') study_id smr_grade_bl (REF=first) visit /PARAM=REF;
       MODEL mtb_l = arm wks cxr_bilat afr_strat cav_bl dtd_bl female smr_grade_bl
                 pcough pcough_bywk dm cigarette_use wks*dm wks*cigarette_use
         /DIST=binomial LINK=logit TYPE3;
       REPEATED SUBJECT=study_id / WITHIN=visit TYPE=ar;
RUN;
```

R code for final GEE model

I chose not to use R to fit my final model, but have included code below for those interested.

At least three R packages may be used to fit GEE models: gee, geepack, and yags. As my dataset included missing observations, I found geepack and yags most useful. These packages permit specification of the sequence of observations when rows with missing data have been omitted. A limitation of geepack is that it may crash while fitting an unstructured correlation structure with lots of covariates; yags does not appear to be similarly affected.

I installed yags within R by using the command

```
install.packages("yags", repos="http://R-Forge.R-project.org")
```

Fitting the model using geepack

```
fit.final <- geeglm(formula = abs(mtb_l - 1) ~ arm + afr_strat + smr_grade_1 +
    smr_grade_3 + smr_grade_4 + dtd_bl + cav_bl + female + pcough +
    cxr_bilat + cigarette_use + dm + weeks_on_tx + pcough_bywk +
    weeks_on_tx * dm + weeks_on_tx * cigarette_use, family=binomial, id=study_id,
    waves=visit, data=thesis, corstr="ar1")
```

Fitting the model using yags

```
fit.final <- yags(abs(mtb_l - 1) ~ arm + afr_strat + smr_grade_1 + smr_grade_3 +
    smr_grade_4 + dtd_bl + cav_bl + female + pcough + cxr_bilat + cigarette_use +
    dm + weeks_on_tx + pcough_bywk + weeks_on_tx*dm + weeks_on_tx*cigarette_use,
    id=study_id, cor.met=visit, family=binomial(), corstruct="ar1", data=thesis,
    alphainit=0.)
```

REFERENCES

1. Glaziou P, Floyd K, Raviglione M. Global burden and epidemiology of tuberculosis. *Clin Chest Med* 2009;30:621–636.

2. Watt C, Hosseini S, Lönnroth K. The global epidemiology of tuberculosis. In: Schaaf H, Zumla A, eds., *Tuberculosis*, Philadelphia, PA: Saunders. 1 ed., 2009;17–26.

3. Sutherland I, Svandova E, Radhakrishna S. The development of clinical tuberculosis following infection with tubercle bacilli. *Tubercule* 1982;63:255–268.

4. Dye C, Scheele S, Dolin P, et al. Global burden of tuberculosis. *JAMA* 1999;282(7):677–686.

5. Trends in tuberculosis– United States, 2010. *MMWR: Morbidity and Mortality Weekly Report* 2011;60(11):333–363.

6. Lönnroth K, Castro K, Chakaya J, et al. Tuberculosis control and elimination 2010−50: cure, care, and social development. *Lancet* 2010;375:1814–1829.

7. Treatment of tuberculosis. *MMWR: Morbidity and Mortality Weekly Report* 2003; 52(RR-11):1–77.

8. Connolly LE, Edelstein PH, Ramakrishnan L. Why is long-term therapy required to cure tuberculosis? *PLoS Med* 2007;4.

9. Jereb JA, Goldberg SV, Powell K, et al. Recommendations for use of an isoniazid-rifapentine regimen with direct observation to treat latent *Mycobacterium tuberculosis* infection. *MMWR: Morbidity and Mortality Weekly Report* 2011;60(48):1650–1653.

10. Zhang Y. Persistent and dormant tubercle bacilli and latent tuberculosis. *Front Biosci* 2004;9:1136–1156.

11. Boshoff H, Barry CE III. Tuberculosis– metabolism and respiration in the absence of growth. *Nature* 2005;3:70–80.

12. Biketov S, Potapov V, Ganina E, et al. The role of resuscitation promoting factors in pathogenesis and reactivation of *Mycobacterium tuberculosis* during intra-peritoneal infection in mice. *BMC Infect Dis* 2007;7:73–77.

13. Mitchison DA. The search for new sterilizing anti-tuberculosis drugs. *Front Biosci* 2004;9:1059–1072.

14. Mukamolova GV, Turapov O, Malkin J, et al. Resuscitation-promoting factors reveal an occult population of tubercle bacilli in sputum. *Am J Respir Crit Care Med* 2010; 181(2):174–180.

15. Fox W, Ellard GA, Mitchison DA. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council Tuberculosis Units, 1946-1986, with relevant subsequent publications. *Int J Tuberc Lung Dis* 1999;3(10):S231–S279.

16. Johnson J, Hadad D, Boom W, et al. Early and extended early bactericidal activity of levofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2006;10(6):605–612.

17. Rustomjee R, Lienhardt C, Kanyok T. A Phase II study of the sterilising activities of ofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2008;12(2):128–138.

18. Glickman SW, Rasiel EB, Hamilton CD, et al. A portfolio model of drug development for tuberculosis. *Science* 2006;311(5765):1246–1247.

19. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ* 2003;22:151–185.

20. Showalter H, Denny WA. A roadmap for drug discovery and its translation to small molecule agents in clinical development for tuberculosis treatment. *Tuberculosis* 2008; 88(Suppl. 1):S3–S17.

21. Medical Research Council Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment for pulmonary tuberculosis. *BMJ* 1948;2:769–782.

22. Hampton J. Size isn't everything. *Stat Med* 2002;21:2807–2814.

23. Laurenzi M, Ginsberg A, Spigelman M. Challenges associated with current and future TB treatment. *Infect Disord Drug Targets* 2000;7(2):105–119.

24. Balganesh TS, Alzari PM, Cole ST. Rising standards for tuberculosis drug development. *Trends Pharmacol Sci* 2008;29(11):576–581.

25. Katz R. Biomarkers and surrogate markers: an FDA perspective. *NeuroRx* 2004; 1(2):189–195.

26. Perrin F, Lipman M, McHugh T, et al. Biomarkers of treatment response in clinical trials of novel antituberculosis agents. *Lancet Infect Dis* 2007;7:481–490.

27. Wallis RS, Pai M, et al DM. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet* 2010;375(4):1920–1937.

28. Nahid P, Saukkonen J, Mac Kenzie WR, et al. Tuberculosis biomarker and surrogate endpoint research roadmap. *Am J Respir Crit Care Med* 2011;184:972–979.

29. van Zyl-Smit R, Binder A, Meldau R, et al. Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS One* 2011;6.

30. Mitchison DA. Assessment of new sterilizing drugs for treating pulmonary tuberculosis by culture at 2 months. *Am J Resp Crit Care* 1993;147:1062–1063.

31. Burman WJ. Rapid molecular detection of tuberculosis and rifampin resistance. *Am J Respir Crit Care Med* 2003;167(10).

32. Davies GR. Early clinical development of anti-tuberculosis drugs: science, statistics and sterilizing activity. *Tuberculosis* 2010;90(3):171–176.

33. Mac Kenzie WR, Heilig CM, Bozeman L, et al. Geographic differences in time to culture conversion in liquid media: Tuberculosis Trials Consortium Study 28. Culture conversion is delayed in Africa. *PLoS One* 2011;6.

34. International Union Against Tuberculosis and Lung Diseases. Technical guide for sputum examination for tuberculosis by direct microscopy. *Bull IUATLD* 1986;61(2):1–16.

35. Hesseling A, Walzl G, Enarson D, et al. Baseline sputum time to detection predicts month two culture conversion and relapse in non-HIV-infected patients. *Int J Tuberc Lung Dis* 2010;14(5):560–570.

36. Gopi P, Chandrasekaran V, Subramani R, et al. Association of conversion & cure with initial smear grading among new smear positive pulmonary tuberculosis patients treated with Category I regimen. *Ind J Med Res* 2006;123:807–814.

37. Horne DJ, Royce SE, Gooze L, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. *Lancet Infect Dis* 2010; 10:387–394.

38. Mitchison DA, Davies GR. Assessment of the efficacy of new anti-tuberculosis drugs. *Open Infect Dis J* 2008;2:59–76.

39. Davies GR, Brindle R, Khoo SH, et al. Use of nonlinear mixed-effects analysis for improved precision of early pharmacodynamic measures in tuberculosis treatment. *Antimicrob Agents Ch* 2006;50(9):3154–3156.

40. Sloan D, Davies G. Serial sputum colony counting in drug development. In: Donald P, van Helden P, eds., *Antituberculosis Chemotherapy*, Basel, Switzerland: S. Karger AG, vol. 40. 2011;227–234.

41. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010;363(11):1005–1015.

42. Blakemore R, Story E, Helb D, et al. Evaluation of the analytical performance of the Xpert MTB/RIF assay. *J Clin Microbiol* 2010;48(7):2495–2501.

43. Helb D, Jones M, Story E, et al. Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol* 2010;48:229–237.

44. Scott L, McCarthy K, Gous N, et al. A roadmap for drug discovery and its translation to small molecule agents in clinical development for tuberculosis treatment. *PLoS Med* 2011;8.

45. Cha D, Cheng D, Liu M, et al. Analysis of fatty acids in sputum from patients with pulmonary tuberculosis using gas chromatography-mass spectrometry preceded by solid-phase microextraction and post-derivatization on the fiber. *J Chromatogr A* 2009; 1216:1450–1457.

46. Kaal E, Kolk A, Kuijper S, et al. A fast method for the identification of *Mycobacterium tuberculosis* in sputum and cultures based on thermally assisted hydrolysis and methylation followed by gas chromatography-mass spectrometry. *J Chromatogr A* 2009; 1216:6319–6325.

47. Ojha A, Baughn A, Sambandan D, et al. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* 2008;69:164–174.

48. Honeyborne I, McHugh T, Phillips P, et al. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum *Mycobacterium tuberculosis* bacillary load during treatment. *J Clin Microbiol* 2011;49(11):3905–3911.

49. Gadkowski LB, Stout J. Cavitary pulmonary disease. *Clin Microbiol* 2008;21(2):305–333.

50. Barnes PF, Leedom JM, Chan LS, et al. Predictors of short-term prognosis in patients with pulmonary tuberculosis. *J Infect Dis* 1988;158(2):366–371.

51. Perrin F, Woodward N, Phillips P, et al. Radiological cavitation, sputum mycobacterial load and treatment response in pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2010; 14(12):1596–1602.

52. Panjabi R, Comstock G, Golub J. Recurrent tuberculosis and its risk factors: adequately treated patients are still at high risk. *Int J Tuberc Lung Dis* 2007;11(8):828–837.

53. Dooley KE, Chaisson RE. Tuberculosis and diabetes mellitus: convergence of two epidemics. *Lancet Infect Dis* 2009;9:737–746.

54. Güler M, Ünsal E, Dursun B, et al. Factors influencing sputum smear and culture conversion time among patients with new case pulmonary tuberculosis. *Int J Clin Pract* 2006;61(2):231–235.

55. Dooley KE, Tang T, Golub JE, et al. Impact of diabetes mellitus on treatment outcomes of patients with active tuberculosis. *Am J Trop Med Hyg* 2009;80(4):634–639.

56. Kumar A, Deshane J, Crossman D, et al. Heme oxygenase-1-derived carbon monoxide induces the *Mycobacterium tuberculosis* dormancy regulon. *J Biol Chem* 2008; 283(26):18032–18039.

57. Javors M, Hatch J, Lamb R. Cut-off levels for breath carbon monoxide as a marker for cigarette smoking. *Addiction* 2005;100:159–167.

58. Creswell J, Raviglione M, Ottmani S, et al. Tuberculosis and noncommunicable diseases: neglected links and missed opportunities. *Eur Respir J* 2011;37(5):1269–1282.

59. Balganesh TS, Alzari PM, Cole ST. Risk of tuberculosis from exposure to tobacco smoke. *Trends Pharmacol Sci* 2008;29(11):576–581.

60. Chen H, Cowan M, Hasday J, et al. Tobacco smoking inhibits expression of proinflammatory cytokines and activation of IL-1R-associated kinase, p38, and NF-κB in alveolar macrophages stimulated with TLR2 and TLR4 agonists. *J Immun* 2007;179:6097–6106.

61. Slama K, Chiang CY, Enarson D, et al. Tobacco and tuberculosis: a qualitative systemic review and meta-analysis. *Int J Tuberc Lung Dis* 2007;11(10):1049–1061.

62. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44(4):1049–1060.

63. Hin LY, Wang YG. Working-correlation structure identification in generalized estimating equations. *Stat Med* 2009;28:642–658.

64. Hilbe JM. *Negative binomial regression.* Cambridge, UK: Cambridge University Press, 2011.

65. Davis C, Hyde J, Bangdiwala S, et al. An example of dependencies among variables in a conditional logistic regression. In: Moolgavkar SH, Prentice RL, eds., *Modern statistical methods in chronic disease epidemiology.* SIAM Institute for Mathematics and Society, vol. 6898, 140–147.

66. Zack M, Singleton J, Wall K, et al. Collinearity diagnostics using the information matrix, cited January 30th, 2012. SAS macro.

67. Kleinbaum DG, Klein M. *Logistic regression.* Berlin, Germany: Springer, 2010.

68. Rotnitzky A, Jewell NP. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika* 1990;77(3):485–497.

69. Boos DD. On generalized score tests. *Amer Statistician* 1992;46(4):327–333.

70. Dorman SE, Goldberg S, Stout JE, et al. Substitution of rifapentine for rifampin during intensive phase treatment of pulmonary tuberculosis: Study 29 of the Tuberculosis Trials Consortium, 2012. Unpublished manuscript.

71. Shults J, Chaganty N. Analysis of serially correlated data using quasileast squares. *Biometrics* 1998;54(4):1622–1630.

72. Ziegler A, Vens M. Generalized estimating equations: Notes on the choice of the working correlation matrix. *Methods Inf Med* 2010;49:421–425.

73. Hilbe JM. *Logistic Regression Models*. Boca Raton, FL: Chapman and Hall/CRC, 2009.

74. Wang YG, Hin LY. Modeling strategies in longitudinal data analysis: Covariate, variance function and correlation structure selection. *Comp Stat Data Anal* 2010;54:3359–3370.

75. Jang M. *Working correlation selection in generalized estimating equations*. Ph.D. thesis, University of Iowa, 2011. Published by Iowa Research Online: http://ir.uiowa.edu/etd/2719.

76. SAS. QIC goodness of fit statistic for GEE models, cited March 10th, 2012. http://support.sas.com/kb/26/100.html.

77. Banu Rekha V, Balasubramanian R, Swaminathan S, et al. Sputum conversion at the end of intensive phase of Category-1 regimen in the treatment of pulmonary tuberculosis patients with diabetes mellitus or HIV infection: an analysis of risk factors. *Indian J Med Res* 2007;126:452–458.

78. Holtz TH, Sternberg M, Kammerer S, et al. Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. *Ann Intern Med* 2006;144(9):650–659.

79. Palaci M, Dietze R, Hadad DJ, et al. Cavitary disease and quantitative sputum bacillary load in cases of pulmonary tuberculosis. *J Clin Microbiol* 2007;45(12):4064–4066.

80. Visser ME, Stead MC, Walzl G, et al. Baseline predictors of sputum culture conversion in pulmonary tuberculosis: importance of cavities, smoking, time to detection and W-Beijing genotype. *PLoS One* 2012;7.

81. Pheiffer C, Carroll N, Beyers N. Time to detection of *Mycobacterium tuberculosis* in BACTEC systems as a viable alternative to colony counting. *Int J Tuberc Lung Dis* 2008;12(7):792–798.

82. Weiner M, Prihoda TJ, Burman W, et al. Evaluation of time to detection of *Mycobacterium tuberculosis* in broth culture as a determinant for end points in treatment trials. *J Clin Microbiol* 2010;48(12):4370–4376.

83. Maâlej S, Belhaoui N, Bourguiba M, et al. La tuberculose pulmonaire provoque un déséquilibre du diabète. *Presse Med* 2009;38(1).

84. Diwan VK, Thorson A. Sex, gender, and tuberculosis. *Lancet* 1999;353:1000–1001.

85. Abal A, Jayakrishnan B, Parwer S, et al. Effect of cigarette smoking on sputum smear conversion in adults with active pulmonary tuberculosis. *Resp Med* 2005;99:415–420.

86. Telzak E, Fazal B, Pollard C, et al. Factors influencing time to sputum conversion among patients with smear-positive pulmonary tuberculosis. *Clin Infect Dis* 1997;25:666–670.

87. Hales CM, Heilig CM, Borisov A, et al. The association between symptoms and microbiologically-defined response to tuberculosis treatment, 2012. Unpublished manuscript.

Tables

Table 2.1: Potential covariates considered for inclusion in preliminary model

Time-invariant	**Reference**
African/non-African stratum	[33]
Age at study entry	[54, 77]
Bilateral cavitation at baseline	[51, 78, 79]
Cavitary status at baseline	[51, 79, 80]
Days to detection at baseline on MGIT 960	[35, 80–82]
Diabetes status at baseline	[55, 58, 83]
Male/female sex	[84]
History of any cigarette use	[54, 58, 59, 61, 85]
History of prior tuberculosis infection	[86]
HIV-status at baseline	[8, 27]
Productive cough at baseline	[87]
Smear grade at baseline	[36, 77]
Time-varying	
Productive cough by week of therapy	[87]

Table 2.2: Bivariate descriptive statistics for 375 participants at baseline

			Treatment Arm		Geographic Region			
	Overall n=375		**Rifapentine** n=197 (52.5%)		**Rifampin** n=178 (47.5%)		**African** n=202 (53.9%)	**Non-African** n=173 (46.1%)	**P-value**[†]
Median values (SD)									
Age in years	33 (13.8)		31 (14.4)		34 (13.2)		28 (9.2)	42 (15.3)	< 0.001
Days to detection	7.1 (3.5)		7.2 (3.5)		7.1 (3.6)		6.2 (2.9)	8.3 (4.0)	< 0.001
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**P-value**[‡]
Female	129	34.4	61	31.0	68	38.2	68	33.7	61	35.3	0.75
History of smoking	153	40.8	81	41.1	72	40.5	64	31.7	89	51.5	< 0.001
Diabetes	42	11.2	23	11.7	19	10.7	1	0.02	41	1.0	< 0.001
Productive cough	334	89.1	178	90.4	156	87.6	194	96.0	140	80.9	< 0.001
Cavitary disease	258	68.8	136	69.0	122	68.5	136	67.3	122	70.5	0.51
Bilateral involvement	229	61.1	125	63.5	104	58.4	138	68.3	91	52.6	0.002
HIV-positive	42	11.2	18	9.1	24	13.5	37	18.3	5	2.9	< 0.001

All *P*-values provided above describe differences between African and non-African strata

[†] Wilcoxon rank sum test

[‡] Pearson's chi-squared test of homogeneity of proportions

Table 2.3: Results for 11 partially adjusted analyses of associations between culture status and proposed predictors

Predictor	**Adjusted OR**	**95% CI**	**P-value***
Baseline measures of bacillary load			
Days to detection (MGIT 960)	1.15	1.09, 1.21	< 0.001
Low smear grade (1-9 bacilli per 10 HPF)	0.58	0.24, 1.42	
Medium smear grade (1-9 bacilli per HPF)	0.26	0.11, 0.66	< 0.001
High smear grade (> 9 bacilli per HPF)	0.16	0.07, 0.40	
Measures of pulmonary tissue damage			
Bilateral involvement (baseline radiograph)	0.72	0.51, 1.03	0.07
Disease-related symptoms			
Productive cough at baseline	0.25	0.14, 0.46	< 0.001
Productive cough by week of therapy	0.54	0.40, 0.72	< 0.001
Also considered as potential predictors			
History of cigarette use	0.77	0.53, 1.11	0.16
History of prior tuberculosis	1.01	0.51, 2.01	0.97
Diabetes	0.81	0.48, 1.38	0.43
Potential confounders			
Female	1.98	1.39, 2.82	< 0.001
Age	0.99	0.98, 1.00	0.15
HIV-positive at baseline	1.29	0.72, 2.31	0.40

* P-value obtained by generalized score test
Abbreviations: HPF – high power field, MGIT – mycobacterial growth in tube

Table 2.4: Final GEE model

Predictor	**Adjusted OR**	**95% CI**	***P*-value***
Weeks on therapy	2.02	1.81, 2.26	< 0.001
Treatment arm	1.32	0.91, 1.91	0.15
African stratum	0.40	0.26, 0.62	< 0.001
Productive cough at baseline	0.28	0.14, 0.58	< 0.001
Cavitary status at baseline	0.68	0.46, 1.02	0.07
Diabetes	0.23	0.06, 0.97	0.027
History of cigarette use	3.07	1.18, 8.00	0.024
Bilateral involvement at baseline	0.92	0.62, 1.35	0.66
Productive cough by week of therapy	0.58	0.42, 0.81	0.002
Baseline days to detection (MGIT 960)	1.09	1.02, 1.16	0.016
Female	1.66	1.11, 2.50	0.016
Weeks on therapy × diabetes			0.013
Week 2	0.43	0.16, 1.15	
4	0.81	0.41, 1.58	
6	1.50	0.72, 3.13	
8	2.80	0.93, 8.43	
Weeks on therapy × cigarette use			0.002
Week 2	1.92	0.96, 3.87	
4	1.20	0.74, 1.96	
6	0.75	0.50, 1.13	
8	0.47	0.28, 0.80	
Smear grade at baseline			0.002
Low: 1-9 bacilli per 10 HPF	0.79	0.32, 1.95	
Medium: 1-9 bacilli per HPF	0.36	0.14, 0.88	
High: > 9 bacilli per HPF	0.32	0.13, 0.80	

* *P*-value obtained from generalized score test
Abbreviations: HPF – high power field, MGIT – mycobacterial growth in tube

Figures

Figure 2.1: Identification of analytic set

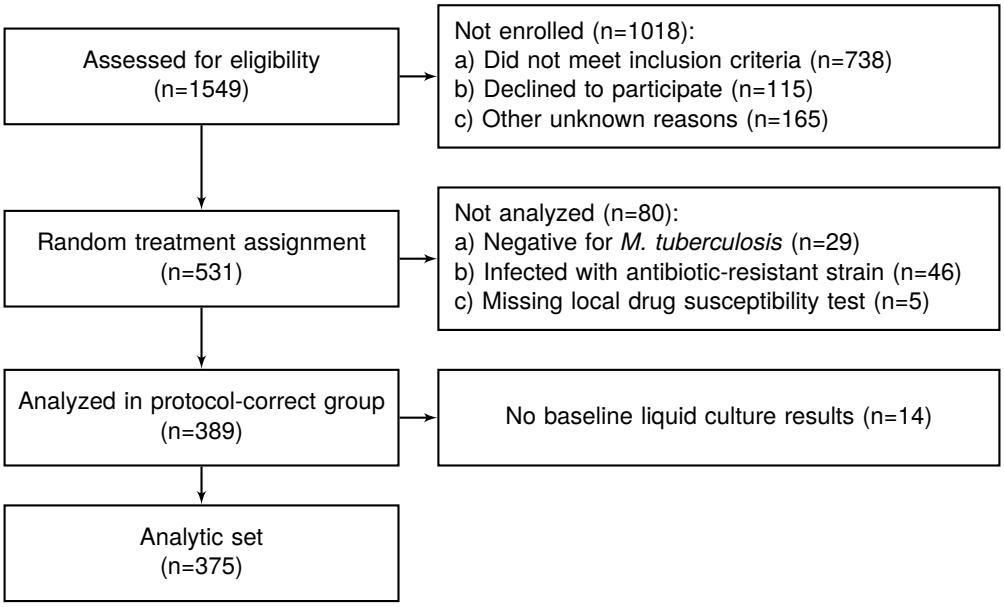
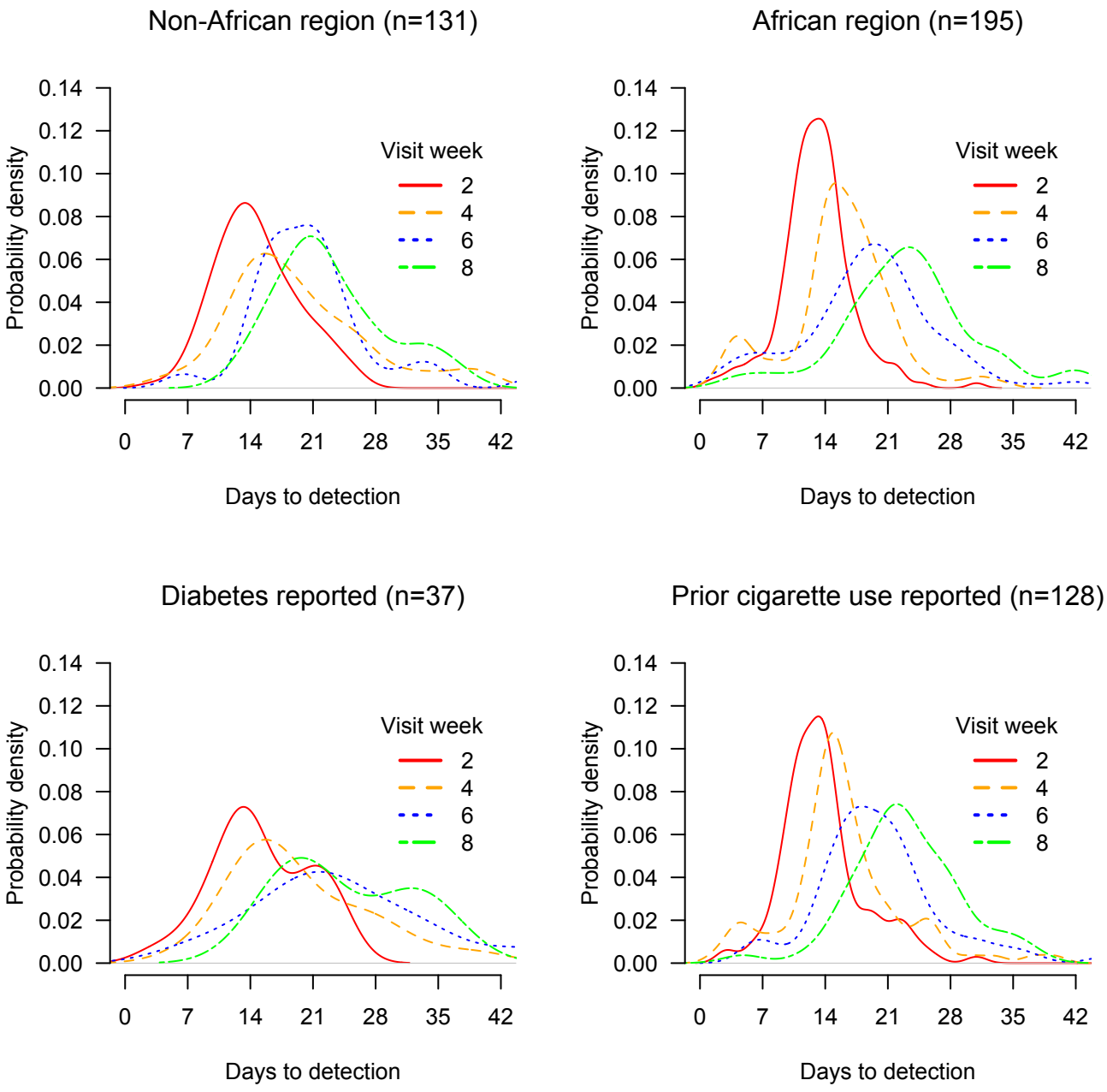
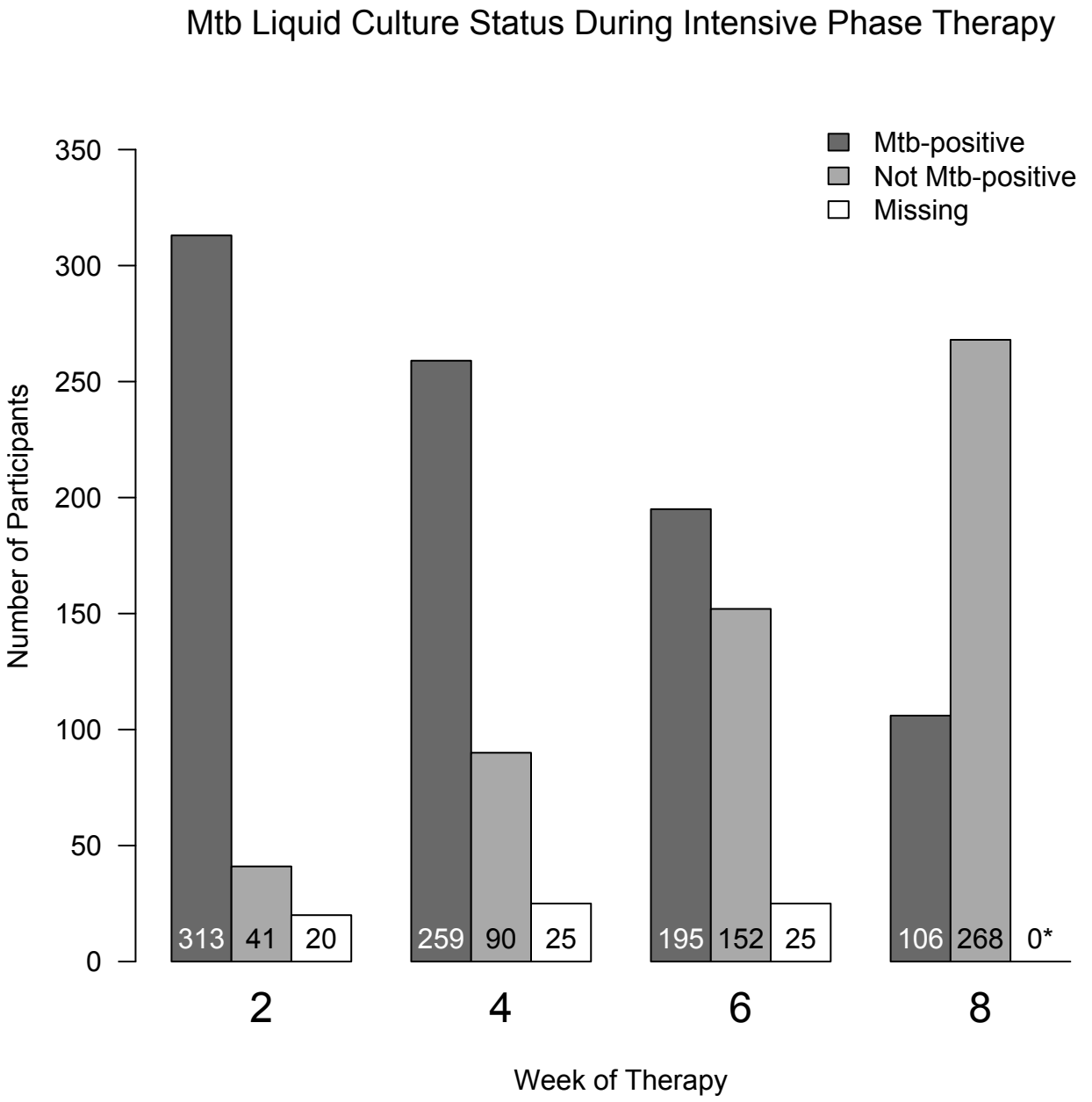


Figure 2.2: Probability densities for days to detection at weeks 2, 4, 6, and 8 of intensive phase therapy as measured among 326 protocol-correct Study 29 participants with detectable Mtb in liquid media



The proportion of missing culture results by each week of therapy may be seen in Figure 2.3. A total of 49 protocol-correct participants had baseline values for days to detection (DTD) using the MGIT 960 system, but did not have detectable Mtb at weeks 2, 4, 6, and 8 of therapy.

Figure 2.3: Culture status by week of therapy among 375 protocol-correct (in liquid media) Study 29 participants



*Although a week 8 culture was unavailable for one protocol-correct participant, two cultures at end of intensive phase therapy were available.

Figure 2.4: Visual comparison of odds ratio estimates and relative precision for all terms in final GEE model