Machine-learning (ML) reveals that *Mycobacterium tuberculosis* genotypes and anatomic disease site impacts drug resistance and disease transmission among patients with proven extra-pulmonary tuberculosis

**CURRENT STATUS:** UNDER REVISION

*BMC Infectious Diseases  BMC Series*

Doctor Busizwe Sibandze (Former Corresponding Author)  
National TB Control Program, Ministry of Health  
*ORCiD:* [https://orcid.org/0000-0002-1331-3124](https://orcid.org/0000-0002-1331-3124)*

Beki Themba Magazi  
MSD (Pty) Ltd

Lesibana Anthony Malinga  
South African Medical Research Council

Nontuthuko Excellent Maningi  
University of Pretoria

Bong Akee Shey  
University of Pretoria

Jotam G Pasipanodya (New Corresponding Author)  
*Corresponding Author*  
JOTAM.PASIPANODYA@ttuhsc.edu

Nontombi N Mbelle  
University of Pretoria

**DOI:**  
10.21203/rs.2.14049/v2

**SUBJECT AREAS**  
*Infectious Diseases*

**KEYWORDS**  
*Stochastic gradient boosting, spoligotypes, number needed to screen, attributable risk*
Abstract
Background There is a general dearth of information on extrapulmonary tuberculosis (EPTB). We investigated Mycobacterium tuberculosis (Mtb) drug resistance and transmission patterns in EPTB patients treated in the Tshwane metropolitan, South Africa. Methods Consecutive Mtb culture-positive non-pulmonary samples from unique EPTB patients underwent further mycobacterial genotyping and were assigned to phylogenetic lineages and transmission clusters based on spoligotypes. MTBDR plus assay was used to test for isoniazid and rifampin susceptibility. ML algorithms were used to identify clinically meaningful patterns in data. We computed odds ratio (OR), attributable risk (AR) and corresponding 95% confidence intervals (CI). Results Of the 70 isolates examined, the largest cluster comprised of 25 (36%) Mtb strains that belonged to the East Asian lineage. East Asian lineage was significantly more likely to occur within chains of transmission when compared to the Euro-American and East-African Indian lineages: OR= 10.11 (95% CI: 1.56-116). Lymphadenitis, meningitis and skin TB, were significantly more likely to be associated with drug resistance: OR=12.69 (95% CI: 1.82-141.60) and AR = 0.25 (95% CI: 0.06-0.43) when compared with other EPTB sites, which suggests that poor rifampin penetration might be a contributing factor.
Conclusions Majority of Mtb strains circulating in the Tshwane metropolis belongs to East Asian, Euro-American and East-African Indian lineages. Each of these are likely to be clustered, suggesting on-going EPTB transmission. Since 25% of the drug resistance was attributable to sanctuary EPTB sites notorious for poor rifampin penetration, we hypothesize that poor anti-tuberculosis drug dosing might have a role in the resistance.

Background
South Africa has one of the highest TB/HIV incidence rate per capita, with 520 confirmed pulmonary TB (PTB) cases per 100,000 population reported in 2015 [1]. It is estimated that 15% to 20% of all TB notified cases might have disease restricted to extra-pulmonary sites (EPTB), such as meningeal, lymphatic, skin or pericardial space. However, the true proportion of proven TB at such anatomical sites is not well described [2, 3]. In order to meet World Health Organization (WHO) tuberculosis (TB) elimination milestones and targets which are: 10 new TB cases per million people per year by the
year 2035 and target goal of 1 case per million people per year by the year 2050, respectively, South Africa will need to undertake more vigorous TB surveillance and direct more resources towards EPTB efforts. However, there are still misinformed beliefs among some public health practitioners and TB programs that EPTB, including childhood TB, do not constitute public health threats because EPTB is less likely to transmit between persons. It is for these and other reasons that childhood TB was not reportable and therefore not formally captured in public records by many national TB programs until 2012 [4]. The net effect has been to regard all TB lesions; except for those from the bronchus, lung parenchyma and bronchopulmonary lymph nodes, in one obscure category called EPTB and devote even fewer resources to the disease [5-7]. As a result, the *Mtb* genotypes, drug resistance patterns and temporal trends associated with EPTB are not well described in South Africa, or in the Tshwane Municipality[3].

The city of Tshwane, which is in Gauteng province, is the financial hub and most densely populated municipality in South Africa. The Tshwane Metropolitan covers a region of 6368 km$^2$ and is supported by a sophisticated network of >25 directly observed treatment strategy (DOTS)/TB centres and tertiary level teaching facilities that serve a multi-ethnic and diverse population in excess of 3 million people, including migrants from across sub-Saharan Africa (*Figure 1*). This makes Tshwane an ideal place to obtain generalizable findings on transmission dynamics of different *Mtb* genotypes with/out drug resistance within diverse populations. The predominant strains associated with the PTB epidemic in the Gauteng province of South Africa are the globally prevalent, modern and re-imported Lineages 4, particularly the so-called Latin-American (LAM) sublineages and the East Asian Lineage 2, which has the moniker “Beijing” strains [2, 8-11]. Both lineages affect geographically diverse human populations worldwide, have been associated with rapid human-to-human transmission and hence greater propensity for acquired drug resistance [9, 12, 13]. On the other hand, the same cannot be said about the more ancient East-African-Indian Lineage 3, such as EAI1_SOM families which have also been isolated with equal frequency in PTB patients from Gauteng province [8, 10, 14]. Here, we wanted to identify factors that predict drug resistance in EPTB patients in Tshwane. In order to inform policy, we specifically wanted to examine the interaction of drug resistance and TB transmission
among EPTB patients in the Tshwane metropolis.

There are several molecular tools that promote the rapid identification of drug resistance patterns of *Mycobacterium tuberculosis* (*Mtb*) complex directly from clinical specimen; and, allow analyses of molecular clock within and against reference [2, 15]. These genotyping technologies can assign the *Mtb* isolates into distinct clusters or groups based on their relatedness with respect to time (phylogeny), geography and other characteristics, in order to ascertain and compare diseases transmission between groups. *Mtb* genotyping is relatively simple, readily available and now affordable. It includes the analyses of IS6100 DNA fingerprints, large sequence polymorphisms, spacer oligonucleotide typing (spoligotyping), Mycobacterial Interspersed Repetitive Units/Variable-Number Tandem-Repeats (MIRU/VNTR), single nucleotide polymorphism (SNP) and more recently, whole-genome sequencing (WGS). Of these, MIRU/VNTR and spoligotyping are the most readily accessible and widely used in developing countries because both have large global reference databases and computational tools that allow assignment of isolates in the major *Mtb* complex lineages [12, 13, 16-18]. When combined with spatio-temporal data, MIRU/VNTR and spoligotyping are considered the gold standard for identifying and tracking chains of *Mtb* transmission. More importantly, both have become indispensable for tuberculosis (TB) epidemiology studies of drug resistance within communities and across the world [5, 7, 14, 16, 19, 20]. The optimal TB treatment duration for the different anatomic sites, including skin, meninges and pericardial spaces, is unknown [5]. Furthermore, there are very few data on the effect of the standard six-to-nine-month combination therapy on acquired drug resistance and patients’ outcomes for these different anatomic sites constituting EPTB. Here we use spoligotyping to identify TB transmission patterns and to characterize clinical *Mtb* isolates obtained from EPTB sites in patients treated in the Tshwane metropolitan of South Africa.

The goal of this study was to use agnostic machine learning (ML) algorithms to determine if there are clear patterns of the different anatomic sites impacting drug resistance and/or genotypic clustering of *Mtb* isolates in affected individuals. Ensembles of ML, such as least absolute selection shrinkage operator (LASSO), classification regression trees (CART) and Random forests, when coupled with
stochastic gradient boosting allows identification of weak base learners and detection of nonlinear dependencies in data, including in pharmacometrics [21-24]. We hypothesized that the majority of EPTB cases will be clustered, with significant proportions bearing drug resistance, which would indicate high rates of transmission of drug resistance in Tshwane. Alternatively, high drug resistance without relationships to clustering would support the ‘pharmacokinetic variability driven de novo acquired resistance’ hypotheses [5, 7, 25, 26], where inadequate drug exposures at sites of TB infection lead to selection for drug resistant and drug tolerant strains which eventually leads to therapy failure [20, 21]. Here we used a combination of ML to examine patterns that were predictive of drug resistance. We specifically examined whether the different EPTB disease sites, the geographic areas of where the patients came from and time when the patients had disease were also predictive of TB transmission in Tshwane.

Methods

Study design

Eligible consecutive clinical samples from EPTB anatomical sites were submitted for diagnosis confirmation by respective DOTS/TB facilities to the National Health Laboratory Services (NHLS). NHLS is the single and integrated laboratory network that covers all public health facilities in South Africa. In Tshwane, NHLS is affiliated with the Department of Medical Microbiology at the University of Pretoria, where microscopy, culture and drug susceptibility testing (DST) of clinical specimens from within Tshwane DOTS/TB facilities as well as other nearby provinces, including Limpopo and Mpumalanga, are performed (Figure 1). All culture positive isolates identified within the six-month study period between July 1st, 2014 and January 31st, 2015 were eligible for enrollment. Isolates that grew nontuberculous Mycobacteria or Mycobacterium bovis were excluded. Patients’ demographics and clinical data were collected from the specimen request forms. Since the analyzed isolates were de-identified and constituted of routinely collected data, the study was not considered human subject research (UP ethical inquiry 143/2015).

Definition of terms

The report of verified case of tuberculosis (RVCT) nomenclature and approach used by the US Centers
for Disease Control and Prevention (CDC) was used to assign collected specimen isolates to EPTB anatomic sites for comparison purposes (https://www.cdc.gov/tb/programs/rvct/rvct-form.pdf).

According to RVCT, EPTB comprises of pleural, peritoneal, lymphatic, meningeal, genitourinary, laryngeal and unclassified group called ‘other’. We modified the RVCT and replaced laryngeal with skin and added a category ‘disseminated’, given the high incidence of those sites in our cohort. Disseminated TB denotes isolates from blood or bone marrow specimens, while patients with a positive isolate from sputum and an EPTB site were grouped separately and denoted ‘PTB/EPTB’. If isolates were obtained from more than one EPTB sites, only one dominant site was recorded. Number needed to screen (NNS) was defined as the number of people that needed to be screened to prevent one TB transmission event or one drug resistance occurrence, based on the assumption that all drug resistance events in the study were acquired during therapy. Isolates that were resistant only to either rifampicin or isoniazid were categorized as mono-resistant, while those resistant to both were categorized as multi drug resistant (MDR) TB.

Spoligotyping examines 43 unique spacer sequences that are interspaced between repetitive sequences in a specific region of difference within the M. tb genome. The presence or absence of each of the 43 variable spacers generates strain-specific fingerprints. In this study, isolates were considered to be clustered if there was an exact match in all 43 spacers. Cluster name and isolate lineage assignments were made by comparison of fingerprints to international databases: https://www.miru-vntrplus.org/MIRU/index.faces, http://www.pasteur-guadeloupe.fr:8081/SITVIT2/index.jsp[18]. The isolates with unmatched genetic profiles were considered nonclustered or orphan strains. The clustering rate was calculated using the following formula: (nc − c)/n, where nc is the total number of clustered isolates, c is the number of isolate clusters, and n is the total number of isolates. Recent transmission of TB infection was presumed to have occurred when a case had an identical spoligotyping pattern to another case in the cohort during six-month study period.

**Laboratory testing methods**

Isolates identified as M.tb were recovered by sub culturing 0.1 ml of the MGIT culture and on
Lowenstein Jansen (LJ) slants to rule out contamination. The slants were placed on their sides, and left at room temperature for 24 hours and thereafter incubated at 37°C for 6 weeks. For genomic DNA extraction from *M. tb*, colonies were swept off from LJ slants and centrifuged at 10,000 g for 15 minutes. The supernatant was discarded and the pellet was re-suspended in 100 ml of sterile distilled water. The specimen was then heat-killed at 95°C for 20 minutes in a water bath. This was followed by sonication for 15 minutes and centrifugation at 13,000 g for 8 minutes. The DNA supernatant was used for spoligotyping and the genotype MTBDRplus line-probe assay while the remainder was stored at -20°C.

The Genotype® MTBDRplus line-probe assay (version 2.0) was performed to determine rifampin and isoniazid susceptibility profiles according to the manufacturer’s instructions (Hain LifeSciences, Nehren, Germany). Briefly, PCR (50 µL/tube; 40-cycle program) was performed using HotStar Taq DNA Polymerase (Qiagen). The PCR products were hybridized following the manufacturer’s instructions. After hybridization, membrane strips were attached to the evaluation sheet, read, and interpreted manually. Spoligotyping was performed and results were analysed using the BioNumerics Software ver. 7.5 (Applied Maths, Kortrijk, Belgium) as previously described by Kamerbeek and colleagues [27]. We assigned each study isolate spoligotype pattern a Spoligotype International Type (SIT) number using the most current international spoligotyping databases comprising of 111,635 clinical isolates [28].

**Classification and regression tree (CART) modeling with the stochastic gradient boosting**

Stochastic gradient boosting was used to identify weak base learners, determine meaningful pairwise interactions and the percent of variance attributable to those interactions, variable important scores for those variables and identify thresholds for decision-making [29-33]. Important variables define the most influential predictors, including both linear and nonlinear rules that appear in the predictive model [33]. We used the methods of the late Leo Breiman [34] a pioneer in machine learning (ML) algorithms and artificial intelligence, and Jerome Friedman [30]. Multivariate adaptive regression spline (MARS) models for binary targets in classification problems implemented in TreeNet version 8.3 software were also used for graphic visualization. Details of the modeling approaches that use these
ML algorithms and tools for pharmacokinetics and pharmacodynamics (PK/PD) analyses, pharmacometrics and for general decision-making purposes in the clinic has been published before and reviewed within [21-23, 25, 35, 36]. Optimal CART for the drug resistance and clustering outcomes were also graphically depicted for illustrative purposes. The variable importance scores from Random forest were used to rank and identify variables most predictive of acquired drug resistance or clustering. CART and MARS in TreeNet were used to identify thresholds for continuous variables applied to clinical decision-making, as we have done in the past [21]. Similarly, both algorithms were used to group categorical variable that were considered similar, based on GINI criteria. Five-fold cross validation was used with all models which included all patients’ clinical characteristics shown in Table 1 and isolates’ spoligotypes. The area under the receiver characteristic, misclassification rates and the F1-statistics were used for model comparison. Parsimony was also used in identify the final models.

**Statistical Analysis**

Output from the gradient boosting ML were used to calculate attributable risk (AR), NNS and in multivariate logistic regression models. Newcombe/Wilson scores with continuity correction were used in computing AR 95% confidence intervals (CI) [37], otherwise exact binomial methods of Klopper-Pearson were employed. The STATA (College Station, Texas) and GraphPad software (San Diego, California) were used for statistical analysis. Fisher’s exact test was used to compare proportions, while the Kruskal-Wallis test compared median values between groups. All tests were two-sided and set at an alpha of 5%.

**Results**

Of the 75 unique and consecutive isolates submitted by only 8 out of the 25 DOTS/TB facilities in Tshwane, we excluded from further analysis 5 (7%) isolates because they grew *Mycobacterium bovis*. All excluded isolates were from children <16 years. Of the remaining 70 (93%) isolates, females contributed 28 (40%). The overall proven EPTB incidence was 4.43 (95% CI: 3.72-5.23) per 100,000 population per year in Tshwane (Figure 2A). Even though the ages varied widely from 1 year to 85 years, only 5 (7.14%) samples were from children <16 years (Table 1). Detailed demographic,
clinical and genotyping data in Table 1 show that women were significantly overrepresented among patients with the East-Africa-Indian genotypes or lineage 3. Figure 2B shows that the most frequently encountered proven EPTB disease sites were pleural and lymphatic each accounting for 29% (95% CI: 18-41%), and skin which accounted for 14% (95% CI: 7-25%), while peritoneal and meningeal each accounted for only 7% (95% CI: 2-16%). However, associations between major M. tb lineages, EPTB disease sites or DOTS/TB facilities were not statistically significant.

**Mycobacterium tuberculosis** spoligopatterns clustering and inferred transmission

Majority of Mtb isolates, 57/70 (81%), occurred in clusters that varied in size from 2 to 25 isolates. The largest cluster comprised of 25 (36%) isolates that all belonged to the Beijing clade, which is an East Asian lineage also called lineage 2 (Figure 3A). Mtb isolates from the three major lineages 2, 3 and 4 were in chain of transmission for 98%, 67% and 70% of the isolates, respectively (Figure 3B). Thus, the Beijing strains were significantly more likely to occur within a chain of EPTB transmission when compared to the Euro-American strain: odds ratio (OR)= 10.11 (95% confidence interval [CI] 1.56-116). On the other hand, 12/13 (92%) of unclustered isolates were unnamed orphans in the international spoligotyping databases, while the other lone isolate belonged to the X2 clade, which is of the Euro-American lineage. Table 2 shows that there was no significant association between clustering and variety of demographic and clinical factors, including notably drug resistance and DOTS/TB facilities, in bivariate analyses based on standard statistical tests.

Next, we applied stochastic gradient ML algorithms to identify the important variables that predicted clustering and to determine if there were nonlinear associations that could explain genotypic clustering patterns (Figure 3C-D). The results shown in Figure 3C revealed that specific EPTB disease site and DOTS/TB facilities as well as any drug resistance were ranked important variables and that nonlinear interaction between these accounted for almost 45% of clustering variance. The usual factors described in TB epidemiology such as age were either less prominent or scored zero (in the case of patients’ gender). The pooled isolates from disseminated TB, lymphadenitis, meningitis, EPTB/PTB and skin TB disease site were less likely to be clustered compared to those not from the same TB disease sites: OR=0.23, 95% CI: 0.10-0.99 and the attributable risk (AR) was 0.18 (95% CI:
0.01-0.40). The OR and AR for clustering improved to 0 (95% CI: 0-0.45) and 0.26 (95% CI: 0.10-0.47), respectively, if DOTS/TB facilities were also used in the combination screening (Figure 3D). If these two factors are used as screening tools the overall NNS to prevent transmission of one TB case would be 3.91 (95% CI: 2.14-10.32. The sensitivity for the CART shown in Figure 3D was 0.56 (95% 0.43-0.68), while the positive predictive value was 0.74 (95% CI 0.60-0.85). However, both the specificity and negative predictive values were poor. Nonetheless, when combined these data show that disease site and DOTS/TB facilities; i.e., geographic information systems, can be used in combination with isolates genotypes to identify situations where TB transmission is taking place, even for paucibacillary diseases such as EPTB.

**Predictors of EPTB drug resistance**

Majority, 59/70 (84%), of isolates were susceptible to both rifampin and isoniazid, while 2 (3%) isolates were MDR-TB. However, rifampin resistance was observed in disproportionately large proportions of isolates, 8/70 (11%), which is rather unusual, since isoniazid mono-resistance was observed in only 3/70 (4%) isolates (Figure 2C). Nonetheless, Table 3 shows that there was no association between drug resistance and most demographic and clinical factors examined, including clustering (p=0.419) or Mtb spoligotypes (p=0.737) for any resistance, based on straightforward frequentist tests. The exception was between rifampin resistance and disease site: p=0.036.

ML analyses revealed the differential impact of the interactions between disease site and Mtb genotypes on any drug resistance and especially MDR-TB/rifampin mono-resistance (Figure 4A/B). Firstly, disease sites characterized by sanctuary states, i.e., lymphadenitis, meningitis and skin TB, were significantly more likely to associate with any drug resistance: OR=12.69 (95% CI: 1.82-141.60) and AR=0.25 (95% CI: 0.06-0.43), when compared to EPTB in other sites. Secondly, with regards to MDR-TB/Rifampin mono-resistance the top predictor was lymphadenitis and skin TB disease sites, which means that meningitis was excluded. This is not surprising since rifampin does not penetrate the blood-brain barrier well and the current doses given are so low that virtually none gets into cerebrospinal spaces. The sensitivity and specificity of using disease site as proxy to identify isolates likely to MDR-TB/rifampin resistant are 1.00 (95% CI: 0.68-1.00) and 0.64 (95% CI: 0.52-0.75),
respectively. When information about likely \textit{Mtb} genotypes is added as shown in Figure 4D, the specificity improves to 0.84 (95% CI: 0.71-0.92). These data suggest that for every four patients (95% CI: 2.11-10.64) with TB lymphadenitis or skin we would expect one or more to have MDR-TB/Rifampin resistance when compared to those with TB in other sites. This means that screening with drug susceptibility tests and changing the treatment regimens would prevent therapy failure and further transmission of drug resistant TB.

Discussion
The study focused on characterizing clinical \textit{Mtb} isolates in real-world settings and hence has limitations related to such observational studies, including inadequate sample size, imprecise and some missing information (in this case HIV infection). First, we only used spoligotypes to assign clusters, which limits and biases the clustering resolution and might potentially over-estimate clustering rates. The second limitation relates to a small sample size and misclassification of EPTB diseases sites, which has notoriously confounded comparison of EPTB incidences between studies [38]. Previous studies have identified disease site-specific risk factors, including those with certain \textit{Mtb} genotypes, drug resistance and meningeal TB which we could not reproduce in our study, since only 5 (7%) meningeal TB and 1 (1%) pericardial TB isolates were enrolled [39-43]. Nonetheless, we used validated RVCT methods for ease of temporal and between studies comparisons. Third, the \textit{Mtb} isolates were not serialized, and information on drug therapy received, TB drug doses and timing of the isolates collected to TB therapy, was not readily available. This made inference and distinction of primary transmitted resistance from acquired resistance in our study difficult [44]. Incomplete medical history on the laboratory request form also made it difficult to determine which isolates were from patients immunosuppressed from HIV or concomitant immunosuppressive agents for rheumatological diseases. Nonetheless, with ML modeling, which is best suited for modeling missing data and highly complex data structure, we were able to demonstrate that routinely collected laboratory and clinical data can be used to screen patients and identify risk groups where acquired drug resistance is most likely to occur. Gradient boosting allows identification of weak predictors, nonlinear relationships and thresholds in the data space,[32] which is like the proverbial “finding a
needle in a haystack”, but in this case one uses giant magnets to find that needle. Sensitivities and specificities >84% are reasonable and actually acceptable, given that the information required for initial screening (i.e., identifying disease site as lymphadenitis, skin TB or meningitis) can be ascertained by history and clinical examination. Moreover, ROC values ~70% with cross validation somehow reassures us that results such as these are likely to be reproducible with similar populations. Unlike most EPTB studies performed at large specialized hospitals [45, 46] our study has minimal referral bias, hence the other strength of this study is that the isolates were from primary DOTS/TB facilities and not from patients treated at tertiary specialized facilities.

There are three notable findings from our study with important public health policy and TB control efforts that target reduction of both disease transmission and drug resistance. The findings are certainly applicable in the Tshwane metropolitan and have potential relevance across similar urban populations in South Africa and across the sub-Saharan African metropoli. The key finding is the hierarchical and nonlinear association between key EPTB disease sites (mainly lymphatic, skin and meningeal) and spoligotypes (mainly to impact both disease transmission and drug resistance. The association between Beijing strains and both TB disease transmission and drug resistance has been well described in South Africa and across the world and the results have been mixed [9-11, 15, 47-49]. Our study demonstrates that those relationships are conditional, complex and characterized by several nonlinear interactions (Figures 3 and 4). For example, two-way interactions between EPTB disease site and another variable explained >20% of the variance in clustering and almost 10% of drug resistance. This means that unless those nonlinear relationships are fully examined, the purported factors driving either transmission or drug resistance will be highly biased or wrong. In fact, for both clustering and drug resistance, the impact of Mtb genotypes is of second-order, which means that bacterial genotypes only acted on some EPTB disease sites and not others. The differential impact of EPTB disease site on any drug resistance (shown in Figure 4C) and on MDR-TB/Rifampin monoresistance (shown in Figure 4D), is actually revealing and consistent with standard PK/PD principles underlying drug resistance emergence [25, 50-52]. PK variability between individuals means that some patients will have faster drug clearances than others when given the same drug.
dose. Therefore, inadequate drug exposures at site of infection, which occurs because of PK variability or suboptimal drug doses or poor drug penetration into protected sites such as meningeal or pericardial spaces, leads to selection of drug resistant or drug tolerant isolates. The selected mutants eventually acquire putative mutations in time. In other words, ADR occurs de novo during therapy primarily because of inadequate dosing or with unoptimized therapy regimens. The WHO recommends the same standardized and uniform therapy regimens and doses used for PTB for EPTB, with the caveat of experts’ opinions that varying longer therapy durations be given for meningeal and bone/joints disease sites [53]. Indeed, these same guidelines are used in Tshwane and as shown in Figures 4C-D is associated with drug resistance in certain EPTB sites such as lymphatic, skin and meningeal site. In this study, the AR for both any resistance and MDR-TB/monoresistance were substantial; 0.25 and 0.64, respectively. The corollary suggestion from this specific finding is that the majority resistance observed in our study are more likely acquired during therapy rather than being ‘pre-existing’ or primary. The NNS for targeted screening among EPTB patients based on disease sites for any resistance was 4 and for MDR-TB/monoresistance 2, which is even more efficient and effective than widely recommended population screenings for active TB in congregate settings or among select risk groups, such as diabetes mellitus or HIV [54]. For example, the NNS HIV infected patients to find one active TB case in low TB incidences places is 25 (ranges 11-144) and high TB incidence places is 10 (ranges 5-22), while that for prisoners was 520 (ranges 69-427) and 43 (ranges 21-123), respectively.

Secondly, even though the proportion of EPTB disease sites were similar to previous observations, the overall incidence of proven EPTB of 4.43 per 100,000 populations was lower than expected. There were 8,034 microbiologically confirmed PTB cases in Tshwane in 2015, an estimated incidence rate of 254 (95% CI: 249-260) cases per 100,000 population [1]. Confirmed PTB status was based on positive GeneXpert MTB/RIF assay, cultures, line probe assays and microscopy smears, which probably overestimated confirmed PTB cases by accepting nontuberculous cases based microscopy smears. Hoogendoorn et al reviewed charts of patients treated and notified for clinical EPTB in the predominantly rural Limpopo province in ten months of 2013 [3]. Of the 336 patients diagnosed, only
57% had good evidence for stated diagnoses. Nonetheless, the overall estimated incidence of clinical EPTB in that study was 27.92 (95% CI: 24.80-31.23) and that for clinical meningeal TB was 2.56 (1.70-3.70) per 100,000 populations per year. Meningeal TB comprised 9.82% (95% CI: 6.86-13.52) of EPTB in Limpopo and 9.04% (95% CI: 6.94-11.54) in Soweto, per year [3, 46]. Our estimates of EPTB incidence in Tshwane is six folds lower than those reported from Limpopo; however, proven meningeal TB comprised 7.14% (2.36-15.89) of cases in Tshwane, suggesting that the meningeal TB proportions were similar between these disparate South African studies. In the US, EPTB as a proportion of total TB cases has been steadily increasing as TB elimination efforts are accelerated and the WHO TB elimination targets getting realized. From 7.6% in 1962 at the peak of the epidemic when TB incidence was 28.6 per 100,000 population, EPTB increased to 15.7% in 1993 with the HIV resurgence and was 30.9% % in 2017 when the reported TB was 2.8 per 100,000 population [55]. Contrary to explanations given by Hoogendoorn and others, we actually hypothesize that with widespread use of methods to identify proven EPTB, the incidence will increase consistent with observations in the US, where majority of EPTB reported are proven TB. We argue that several cases currently reported as clinical EPTB by Hoogendoorn and others in South Africa and elsewhere in low-resources settings could actually be other bacterial infections or due to systemic inflammation from HIV infection.

EPTB is generally paucibacillary in nature which means that usually there are not enough TB bacilli in tissues from which cultures can be obtained; while histology samples are not easy to obtain and therefore not routinely collected. Culture positivity and histology examination of clinical samples, which are the gold standards for confirming EPTB, are notoriously low (about 15% in high TB burden areas) and inconsistent when compared against clinically suspected TB cases. Investments in improved diagnostics to confirm proven EPTB or ML algorithms that are trained on large clinical data to predict EPTB, will not only save lives by reducing unnecessary TB treatments, but will also be cost-effective because of reduction TB diseases transmission and ADR. Both interventions will accelerate meeting TB elimination targets.

Finally, with regards to the heterogeneity of the \textit{Mtb} spoligotypes causing EPTB and the general
predominance of the Beijing clades (lineage 2) and the Euro-American lineage 4 within the Tshwane metropolis, our results are in concordance with the work of others [19]. This is not surprising since the lineages 2 and 4 are thought to be the most successful strains among all the Mtb complex organisms in causing all forms of TB disease, including PTB [13,18, 20]. Previous reports have associated Mtb lineages of Beijing clade with major outbreaks in different parts of the world and was shown to disseminate more rapidly and caused more-severe disease than other strains [21-23]. Moreover, several other epidemiological data suggest that certain M.tb genotypes, such as the W-Beijing genotypes, are more transmissible than others [20-22]. Our study found that the Beijing strains within a chain EPTB transmission was statistically significant when compared to the Euro-American and East African Indian strains which might support the variable virulence hypotheses [23,24].

Conclusion
A majority of Mtb strains circulating in the city of Tshwane metropolis are East Asian (predominantly Beijing clade), Euro-American and East-African Indian lineages, and each of these are likely to be clustered, suggesting on-going transmission of both drug-susceptible and drug-resistant EPTB disease. However, the proportion attributed to transmission was significantly higher with the East Asian lineage compared to the other lineages, which might support the variable virulence hypothesis. On the other hand, the proportion of drug resistance, especially rifampin resistance, attributable to certain sanctuary EPTB sites, including lymph nodes, meninges and skin, was significantly higher, 25% (95% CI: 0.06-0.), when compared with other EPTB sites. This observation suggests that low rifampin exposures, due to poor penetration into those sites or inadequate rifampin doses, significantly contribute to acquired drug resistance, which is also consistent with PK/PD principles of PK variability. Moreover, the significant nonlinear relationship between EPTB sites, Mtb genotypes and drug resistance (particularly MDR-TB/Rifampin mono-resistance) observed, is consistent with prior clinical observations. Together, these data suggest that inadequately treated EPTB is contributing to drug resistance and overall poorer outcomes.

Abbreviations
TB: Tuberculosis, PTB: Pulmonary tuberculosis, EPTB: Extra-pulmonary tuberculosis, DST: drug susceptibility testing, LPA: Line probe assay, INH: isoniazid, MDR-TB: multidrug resistant tuberculosis, RIF: rifampicin, WHO: World Health Organization, NHLS/TAD: National Health Laboratory Services/Tshwane Academic Division, MRC: Medical Research Council.

Declarations

Acknowledgements

The authors wish to thank the employees of the National Health Laboratory Services, Tshwane Academic division, and Dr. Natasha Morris of the Health GIS Centre, South African Medical Research Council for their technical support for this study.

Availability of Data and Materials

The data and materials from this study is stored in the National Health Laboratory Track Care lab database.

Author’s contributions

DB wrote the first draft manuscript, thereafter NM, JGP and BM contributed jointly to writing the revised manuscript. JGP analyzed data and performed the Machine-learning modeling. Clinical aspects of the manuscript were contributed by NM, JGP and BM. Laboratory aspects of the manuscript were contributed by DS, BS, NM, and LM. All authors read and approved the final manuscript.

Competing Interests

All authors declare that they have no competing interests.

Consent for Publication

Not Applicable

Ethical approval

Ethical approval was obtained from the Faculty of Health Science Research Ethics, University of Pretoria with protocol number 143/2015 and preceded experimental work.

References

1. Nanoo A, Izu A, Ismail NA, Ihekweazu C, Abubakar I, Mametja D, Madhi SA:

   Nationwide and regional incidence of microbiologically confirmed pulmonary
tuberculosis in South Africa, 2004-12: a time series analysis. Lancet Infect Dis 2015, 15(9):1066-1076.

2. Dheda K, Gumbo T, Maartens G, Dooley KE, Mcnerney R, Murray M, Furin J, Nardell EA, London L, Lessem E et al: The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. Lancet Respir Med 2017.

3. Hoogendoorn JC, Ranoto L, Muditambi N, Railton J, Maswanganyi M, Struthers HE, McIntyre JA, Peters RPH: Reduction in extrapulmonary tuberculosis in context of antiretroviral therapy scale-up in rural South Africa. Epidemiol Infect 2017, 145(12):2500-2509.

4. Marais BJ, Graham SM, Maeurer M, Zumla A: Progress and challenges in childhood tuberculosis. Lancet Infect Dis 2013, 13(4):287-289.

5. Pusch T, Pasipanodya JG, Hall RG, 2nd, Gumbo T: Therapy duration and long-term outcomes in extra-pulmonary tuberculosis. BMC Infect Dis 2014, 14:115.

6. Pasipanodya JG, Nuernberger E, Romero K, Hanna D, Gumbo T: Systematic analysis of hollow fiber model of tuberculosis experiments. Clin Infect Dis 2015, 61 Suppl 1:S10-17.

7. Shaw JE, Pasipanodya JG, Gumbo T: Meningeal tuberculosis: high long-term mortality despite standard therapy. Medicine (Baltimore) 2010, 89(3):189-195.

8. Said HM, Kock MM, Ismail NA, Mphahlele M, Baba K, Omar SV, Osman AG, Hoosen AA, Ehlers MM: Molecular characterization and second-line antituberculosis drug resistance patterns of multidrug-resistant Mycobacterium tuberculosis isolates from the northern region of South Africa. J Clin Microbiol 2012, 50(9):2857-2862.

9. Streicher EM, Muller B, Chihota V, Mlambo C, Tait M, Pillay M, Trollip A, Hoek KG,
Sirgel FA, Gey van Pittius NC et al: Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. Infect Genet Evol 2012, 12(4):686-694.

10. Stavrum R, Mphahlele M, Ovres K, Muthivhi T, Fourie PB, Weyer K, Grewal HM: High diversity of Mycobacterium tuberculosis genotypes in South Africa and preponderance of mixed infections among ST53 isolates. J Clin Microbiol 2009, 47(6):1848-1856.

11. Maningi NE, Daum LT, Rodriguez JD, Said HM, Peters RPH, Sekyere JO, Fischer GW, Chambers JP, Fourie PB: Multi- and Extensively Drug Resistant Mycobacterium tuberculosis in South Africa: a Molecular Analysis of Historical Isolates. J Clin Microbiol 2018, 56(5).

12. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW et al: High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 2008, 6(12):e311.

13. Wirth T, Hildebrand F, Allix-Beguec C, Wolbeling F, Kubica T, Kremer K, van Soolingen D, Rusch-Gerdes S, Locht C, Brisse S et al: Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog 2008, 4(9):e1000160.

14. Pasipanodya JG, Moonan PK, Vecino E, Miller TL, Fernandez M, Slocum P, Drewyer G, Weis SE: Allopatric tuberculosis host-pathogen relationships are associated with greater pulmonary impairment. Infect Genet Evol 2013, 16:433-440.

15. Meehan CJ, Moris P, Kohl TA, Pecerska J, Akter S, Merker M, Utpatel C, Beckert P, Gehre F, Lempens P et al: The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology. EBioMedicine 2018, 37:410-416.
16. Tulu B, Ameni G: Spoligotyping based genetic diversity of Mycobacterium tuberculosis in Ethiopia: a systematic review. *BMC Infect Dis* 2018, 18(1):140.

17. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC et al: Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 2006, 103(8):2869-2873.

18. Allix-Beguec C, Harmsen D, Weniger T, Supply P, Niemann S: Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis complex isolates*. *J Clin Microbiol* 2008, 46(8):2692-2699.

19. Juma SP, Maro A, Pholwat S, Mpagama SG, Gratz J, Liyoyo A, Houpt ER, Kibiki GS, Mmbaga BT, Heysell SK: Underestimated pyrazinamide resistance may compromise outcomes of pyrazinamide containing regimens for treatment of drug susceptible and multi-drug-resistant tuberculosis in Tanzania. *BMC Infect Dis* 2019, 19(1):129.

20. Magombedze G, Garira W, Mwenje E: Modelling the human immune response mechanisms to mycobacterium tuberculosis infection in the lungs. *Math Biosci Eng* 2006, 3(4):661-682.

21. Swaminathan S, Pasipanodya JG, Ramachandran G, Hemanth Kumar AK, Srivastava S, Deshpande D, Nuermberger E, Gumbo T: Drug Concentration Thresholds Predictive of Therapy Failure and Death in Children With Tuberculosis: Bread Crumb Trails in Random Forests. *Clin Infect Dis* 2016, 63(suppl 3):S63-S74.

22. Pasipanodya JG, Smythe W, Merle CS, Olliaro PL, Deshpande D, Magombedze G, McIlreron H, Gumbo T: Artificial intelligence-derived 3-Way Concentration-dependent Antagonism of Gatifloxacin, Pyrazinamide, and Rifampicin During
**Treatment of Pulmonary Tuberculosis.** *Clin Infect Dis* 2018, **67**(suppl_3):S284-S292.

23. Beki T. Magazi SK, Robert Badal, Tawanda Gumbo, Jotam G. Pasipanodya: **AI Molecular Characterization and Prediction in Carbapenem Resistant Enterobacteriaceae from Intra-abdominal Infections in the SMART Study.** *Journal of Antimicrobial and Chemotherapy* 2018, *In Review*.

24. Gestal M, Dedloff, MR., Torres-Sangiao, E: **Computational Health Engineering Applied to Model Infectious Diseases and Antimicrobial Resistance Spread.** *Applied Sciences* 2019, **9**(12):2486.

25. Pasipanodya JG, Gumbo T: **A new evolutionary and pharmacokinetic-pharmacodynamic scenario for rapid emergence of resistance to single and multiple anti-tuberculosis drugs.** *Curr Opin Pharmacol* 2011, **11**(5):457-463.

26. Srivastava S, Pasipanodya JG, Meek C, Leff R, Gumbo T: **Multidrug-resistant tuberculosis not due to noncompliance but to between-patient pharmacokinetic variability.** *J Infect Dis* 2011, **204**(12):1951-1959.

27. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M et al: **Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology.** *J Clin Microbiol* 1997, **35**(4):907-914.

28. Couvin D, David A, Zozio T, Rastogi N: **Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis genotyping database.** *Infect Genet Evol* 2018.

29. Breiman L, Friedman J, Stone CJ, Olshen RA: **Classification and Regression Trees.** Boca Raton: Chapman and Hall/CRC; 1984.
30. Breiman L: *Statistical Modeling: The Two Cultures* (with comments and a rejoinder by the author). *Statistical Science* 2001, 16(3):199-231.

31. Friedman JH: *Multivariate adaptive regression splines*. *AnnStatist* 1991, 19(1):1-68.

32. Friedman JH: *Stochastic gradient boosting*. In. Stanford University: Department of Statistics; 1999.

33. Friedman J, Popescu, BE: *Predictive learning via rule ensembles*. *Ann Appl Stat* 2008, 2(3):916-954.

34. Breiman L: *Convergence Properties of a Learning Algorithm*. *The Annals of Mathematical Statistics* 1964, 35(4):1819-1822.

35. Pasipanodya J, Gumbo T: *An oracle: antituberculosis pharmacokinetics-pharmacodynamics, clinical correlation, and clinical trial simulations to predict the future*. *Antimicrob Agents Chemother* 2011, 55(1):24-34.

36. Gumbo T, Pasipanodya JG, Romero K, Hanna D, Nuermberger E: *Forecasting accuracy of the hollow fiber model of tuberculosis for clinical therapeutic outcomes*. *Clin Infect Dis* 2015, 61 Suppl 1:S25-31.

37. Newcombe RG: *Interval estimation for the difference between independent proportions: comparison of eleven methods*. *Stat Med* 1998, 17(8):873-890.

38. Kulchavenya E: *Extrapulmonary tuberculosis: are statistical reports accurate?* *Ther Adv Infect Dis* 2014, 2(2):61-70.

39. Caws M, Drobniewski FA: *Molecular techniques in the diagnosis of Mycobacterium tuberculosis and the detection of drug resistance*. *AnnNYAcadSci* 2001, 953:138-145.

40. Thwaites G, Caws M, Chau TT, D'Sa A, Lan NT, Huyen MN, Gagneux S, Anh PT, Tho DQ, Torok E et al: *Relationship between Mycobacterium tuberculosis genotype*
and the clinical phenotype of pulmonary and meningeal tuberculosis. J Clin Microbiol 2008, 46(4):1363-1368.

41. Thwaites GE, Chau TT, Caws M, Phu NH, Chuong LV, Sinh DX, Drobniewski F, White NJ, Parry CM, Farrar JJ: Isoniazid resistance, mycobacterial genotype and outcome in Vietnamese adults with tuberculous meningitis. Int J Tuberc Lung Dis 2002, 6(10):865-871.

42. Vinnard C, Winston CA, Wileyto EP, MacGregor RR, Bisson GP: Multidrug resistant tuberculous meningitis in the United States, 1993-2005. J Infect 2011, 63(3):240-242.

43. Vinnard C, Winston CA, Wileyto EP, MacGregor RR, Bisson GP: Isoniazid-resistant tuberculous meningitis, United States, 1993-2005. Emerg Infect Dis 2011, 17(3):539-542.

44. Dheda K, Lenders L, Magombedze G, Srivastava S, Raj P, Arning E, Ashcraft P, Bottiglieri T, Wainwright H, Pennel T et al: Drug-Penetration Gradients Associated with Acquired Drug Resistance in Patients with Tuberculosis. Am J Respir Crit Care Med 2018, 198(9):1208-1219.

45. Edginton ME, Wong ML, Phofa R, Mahlaba D, Hodkinson HJ: Tuberculosis at Chris Hani Baragwanath Hospital: numbers of patients diagnosed and outcomes of referrals to district clinics. Int J Tuberc Lung Dis 2005, 9(4):398-402.

46. Karstaedt AS: Extrapulmonary tuberculosis among adults: experience at Chris Hani Baragwanath Academic Hospital, Johannesburg, South Africa. S Afr Med J 2013, 104(1):22-24.

47. Maguga-Phasha NTC, Munyai NS, Mashinya F, Makgatho ME, Mbagiorgu EF: Genetic diversity and distribution of Mycobacterium tuberculosis genotypes in Limpopo, South Africa. BMC Infect Dis 2017, 17(1):764.
48. Nicol MP, Sola C, February B, Rastogi N, Steyn L, Wilkinson RJ: Distribution of strain families of Mycobacterium tuberculosis causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. *J Clin Microbiol* 2005, 43(11):5779-5781.

49. Middelkoop K, Bekker LG, Mathema B, Shashkina E, Kurepina N, Whitelaw A, Fallows D, Morrow C, Kreiswirth B, Kaplan G et al: Molecular epidemiology of *Mycobacterium tuberculosis* in a South African community with high HIV prevalence. *J Infect Dis* 2009, 200(8):1207-1211.

50. Srivastava S, Musuka S, Sherman C, Meek C, Leff R, Gumbo T: Efflux-pump-derived multiple drug resistance to ethambutol monotherapy in *Mycobacterium tuberculosis* and the pharmacokinetics and pharmacodynamics of ethambutol. *J Infect Dis* 2010, 201(8):1225-1231.

51. Schmalstieg AM, Srivastava S, Belkaya S, Deshpande D, Meek C, Leff R, van Oers NS, Gumbo T: The antibiotic resistance arrow of time: efflux pump induction is a general first step in the evolution of mycobacterial drug resistance. *Antimicrob Agents Chemother* 2012, 56(9):4806-4815.

52. Pasipanodya JG, Srivastava S, Gumbo T: Acquired drug resistance because of pharmacokinetic variability in a young child with tuberculosis. *Pediatr Infect Dis J* 2014, 33(11):1205.

53. World Health O: *Treatment of tuberculosis: Guidelines. Fourth edition.* In., vol. WHO/HTM/TB/2009.420. Geneva: World Health Organization; 2010.

54. Kranzer K, Afnan-Holmes H, Tomlin K, Golub JE, Shapiro AE, Schaap A, Corbett EL, Lonnroth K, Glynn JR: The benefits to communities and individuals of screening for active tuberculosis disease: a systematic review. *Int J Tuberc Lung Dis* 2013, 17(4):432-446.
55. Peto HM, Pratt RH, Harrington TA, LoBue PA, Armstrong LR: Epidemiology of extrapulmonary tuberculosis in the United States, 1993-2006. Clin Infect Dis 2009, 49(9):1350-1357.

Tables
Due to technical limitations, Tables 1 - 3 are only available for download from the Supplementary Files section.

Figures
Geographic location of tuberculosis (TB) services and directly observed treatment strategy
(DOTS) centers in Tshwane municipality of South Africa.
Population estimates of extrapulmonary tuberculosis (EPTB) and proportion with drug resistance in Tshwane by age group and sex in 2015: Figure 2A shows that the estimated EPTB incidence stratified by age and gender. As shown the estimates in females was 3.54 (95% CI: 2.68-4.60), while that in males was 5.31 (95% CI: 4.24-6.58) per 100,000 populations for the year 2015. Figure 2B shows the proportion of total isolates (N=70) by anatomic EPTB and within each category the percent of isolates with resistance to either rifampin or isoniazid or both. As shown, none of the isolates from peritoneal specimen, disseminated (i.e., blood or bone marrow) and other specimen samples were drug resistant. Figure 2C stratifies drug resistance by age, and as shown one out of the 5 isolates from children were drug resistant, and that same isolate was also rifampin resistance.
Clustering and chains of Mycobacterium tuberculosis transmission: The number of clusters and the sizes of each cluster are shown in Figure 3A, while the proportion of patients from each major genotype lineages (2, 3 and 4) in a chain of transmission are depicted in Figure 3B (there were no isolates from lineage 1 enrolled in study). Variable importance scores and proportion of the variance explained by interactions between variables were obtained from stochastic gradient modeling of between 200 and 2000 classification and regression trees (CART) are shown in Figure 3C, while the optimal and sample tree from those models is shown in Figure 3D. Disease site was the most important variable at the apex with 100%, while DOTS/TB Facility was second with 92% relative to disease site. However, between variables interactions explained 21% of the variance for disease site and 19% for DOTS/TB Facility (Figure 2C) which means that there are important nonlinear
interactions accounting clustering variance. Figure 3D shows disease site and DOTS/TB Facility interactions significantly influence clustering, even though each individual variable was not statistically significant in Table 2 based on Fischer’s exact test. As shown in, isolates from disseminated diseases, lymph nodes, meninges, EPTB/PTB and skin were significantly less to be clustered; 32/43 (74%) versus 25/27 (93%), when compared to the rest of disease site. The receiver operating characteristics curve (ROC) for this single node is 0.744 (95% confidence interval [CI] 0.590-0.991). The model is reproducible as demonstrated by the test ROC of 0.688 and error rate of <3% on the training model.

![Image of Figure 3D]

Figure 3D

Predictors of drug resistance in Mycobacterium tuberculosis isolates from extra-pulmonary
sites: The variable importance scores and proportion of the variance explained by interactions between the variables that were obtained from stochastic gradient modeling for any drug resistance are shown in Figure 4A, while those for MDR-TB/Rifampin monoresistance are shown in Figure 4B. Multivariate adaptive regression trees (MARS) for binary outcomes with two-way interactions detection were made in the TreeNet software. The optimal representative classification and regression trees (CART) are shown in Figure 4C for any resistance and in Figure 4D for MDR-TB/Rifampin monoresistance. The primary node (disease site) for any drug in Figure 4C is almost identical to that for MDR-TB/Rifampin monoresistance in Figure 4D, the difference being addition of meninges to the former group. The sensitivity for both is 0.72 (95% CI: 0.56-0.84). However, positive predictive value for the former is 0.44 (95% 0.32-0.57) and for the latter is 0.36 (95% 0.25-0.48). The MDR-TB/Rifampin monoresistance group necessarily excludes the three isoniazid monoresistance isolates, hence the overall number of isolates analyzed in Figures 4C/D are 67 and not 70.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Table 1.pdf
Table 2.pdf
Table 3.pdf