Causes and consequences of adult sepsis in Blantyre, Malawi

-

Thesis submitted in accordance with the requirements of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy by Joseph Michael Lewis

August 2019

Contents

List of Tables

List of Figures

Preface

Placeholder

Chapter 1

Introduction

Placeholder

1.1 Chapter Overview

1.2 Sepsis in sub-Saharan Africa

1.2.1 Search strategy

1.2.2 Defining sepsis

1.2.3 Applicability of sepsis-3 definitions in sub-Saharan Africa

1.2.4 Sepsis epidemiology in sub-Sahara Africa

1.2.4.1 Incidence

1.2.4.2 Risk factors: the sepsis population in sub-Saharan Africa

1.2.4.3 Outcomes

1.2.5 Sepsis aetiology in sub-Saharan Africa

1.2.5.1 Bacterial zoonoses, Rickettsioses and arboviruses

1.2.5.2 HIV opportunistic infections: PCP, histoplasmosis and cryptococcal disease

1.2.6 Sepsis management

1.2.6.1 Early goal directed therapy

1.2.6.2 Evidence to guide antimicrobial therapy in sSA

1.2.6.3 Evidence to guide intravenous fluid therapy in sub-Saharan Africa

1.3 ESBL-E in sub-Saharan Africa

1.3.1 Search strategy

1.3.2 Introduction: definition and classification of ESBL-E

1.3.3 Global molecular epidemiology of ESBL-E: an overview

1.3.3.1 1980s-1990s: First identificatiom of ESBL in nosocomial pathogens

1.3.3.2 1990s-2010s: Emergence and globalisation of CTX-M

Chapter 2

Methods

Placeholder

2.1 Chapter Overview

2.2 Study site

2.2.1 Malawi

2.2.2 Queen Elizabeth Central Hospital

2.2.3 Participating Laboratories

2.2.3.1 Malawi-Liverpool-Wellcome Clinical Research Programme

2.2.3.2 Malawi College of Medicine Tuberculosis Laboratory

2.2.3.3 Wellcome Trust Sanger Institute

2.3 Clinical Study

2.3.1 Entry Criteria

2.3.2 Study Visits and Patient Sampling

2.3.2.1 Enrollment assessment and first six hours

2.3.2.2 Subsequent visits

2.3.2.3 Blood, urine, and stool, sputum and CSF collection

2.3.2.4 Imaging: chest x-ray and ultrasound scanning

2.3.3 Outcomes and sample size calculations

2.4 Diagnostic Laboratory Procedures

2.4.1 Point of care diagnostics

2.4.2 Laboratory diagnostics

2.4.2.1 Haematology and biochemistry

2.4.2.2 Aerobic blood and CSF culture

2.4.2.3 Mycobacterial blood culture

Chapter 3

A clinical and microbiological description of sepsis in Blantyre, Malawi

Placeholder

3.1 Chapter overview

3.2 Introduction and chapter aims

3.3 Methods

3.4 Results

3.4.1 Study population

3.4.2 Baseline characteristics

3.4.3 Admission physiology and laboratory investigations

3.4.4 Aetiology

3.4.5 Treatment

3.4.6 Outcome

3.4.7 Determinants of mortality

3.5 Discussion

3.5.1 Demographics and outcome: significant longer-term mortality

3.5.2 Aetiology: TB dominates as a cause of sepsis

3.5.3 Determinants of mortality

3.5.4 Limitations

3.6 Conclusions and further work

Chapter 4

Exploratory modelling of sepsis outcome

4.1 Introduction and chapter aims

and to develop models to attempt to understand the causal effect of interventions delivered to patients presenting with sepsis.

The second aim - models to understand the causal effect of interventions delivered to patients with sepsis - presents conceptual and technical difficulties, however. There are a number of standard modelling approaches in the biomedical literature when putative associations between predictor variables and mortality are to be identified. The usual approach consists of selecting variables using some criteria as variables to be included as predictors in a regression model, and identified associations are interpreted as the independent effect of the included variables. There are two common problems with this approach. Firstly, commonly used variable selection strategies have the possibility of introducing significant bias, if they use associations within the data to guide inclusion of variables e.g. bivariate associations or stepwise variable inclusion using statistical significance (or other) thresholds. This is because the statistics used to test the parameters (and generate confidence intervals around effect sizes etc.) are based on an assumption that a single hypothesis is being tested, an assumption which is violated by the stepwise model building process. It can be shown that standard errors are too small, that p-values are biased towards zero and parameter estimates biased away from zero [1]. But selecting variables to be included in a regression model is a difficult problem with no consensus on an ideal solution; *a priori* selection of variables for theoretical reasons is likely ideal, but this becomes difficult when there are a large number of potentially important predictors. This is because including more predictor variables - though it may

decrease bias in the estimates of the model - increases the variance of the predicted values, the so-called bias-variance trade off. Dimensionality reduction techniques (such as principal components analysis) or shrinkage methods (lasso or ridge regression) have been suggested as alternative predictor variable selection techniques[1–3]. A further problem in modelling mortality in studies of sick inpatients is collinearity, where some predictor variables can be predicted with high accuracy by other predictor variables. For example, shocked patients are likely to have elevated lactate, low blood pressure, low bicarbonate, and high heart rate and so parameter estimates become very large when these are all entered a regression model together. An advantage of principal-components type dimensionality reduction is that they can solve this problem by generation new coordinate systems that are constrained to be orthogonal.

Secondly, even if a regression model is correctly specified in terms of predictor variables, correct interpretation of predictor effects is often difficult or impossible without a clear hypothesised causal structure. For example, consider a hypothesised causal structure of death in sepsis in Figure 4.1, which I express as a directed acyclic graph (DAG); nodes represent collections of variables which theoretically specify host status (age, sex, immune status including HIV status and CD4 cell count), infection type (e.g. causative pathogen, site), disease severity (e.g physiological variables quantifying shock, hypoxia etc.), therapies administered, and outcome. Arrows (called edges in the DAG framework) show causality: host status influences infection (e.g. TB is more common in HIV) and severity (patients with advanced HIV may have more severe infection), for example, and therapies administered is likely to be influenced by disease severity (perhaps sicker patients receive antimicrobials more quickly), host status (clinicians are likely to administer different therapies to HIV-infected patients), and infection type. A standard analysis of sepsis would construct a predictive multivariable model for death by including factors which the analyst felt likely to be associated with mortality, which would usually include HIV status, CD4 cell count, physiologic variables (such as presence of shock) and infection variables (e.g. presence of bloodstream infection [BSI]). The effects of the predictor variables are often then interpreted as the independent effect of the included predictors, after controlling for all others; however, this may not be the case.

For example, severity is at least in part a mediator of the effect of HIV on outcome, so the interpretation of the coefficient of HIV in such a model is the residual effect of HIV once disease severity is accounted for. It is likely that there are direct effects of host and infection factors on outcome (dotted edges in Figure 4.1, not least because measured variables in a study are unlikely to wholly quantify disease severity, but if not then controlling for disease severity will completely remove the effect of HIV status on mortality, which may not be the analysts intention, or interpretation of parameters. This has been called the “Table 2 fallacy.”[4] It is important therefore to clearly define the effect that is being sought from an analysis (e.g. the effect of HIV status on mortality) and to ascertain which factors need to be

controlled for based on this. It may be that a number of different models are necessary to estimate parameters of interest, if more than one parameter is of interest. The causal inference framework provides tools to do this using DAGs[5], and the *dagitty* package in R[6] automates this framework so, when provided with a DAG, it can output the variables that must be conditioned upon to estimate the causal effect of an exposure on an outcome. In this chapter, therefore, I am clear that the aim of the analysis is to provide an estimate of the effect of treatments administered on mortality; the class of antimicrobial administered (antibacterial, antifungal, ant-TB or antimalarial) as well as the time-to-antimicrobial for different classes, and the volumes of intravenous fluid administered. This will inform the overarching aim of the thesis - to develop novel antimicrobial strategies for sepsis in sSA to improve outcomes.

4.2 Methods

Assuming the causal model in 4.1, an estimation of the effect of administered treatment will require correcting for (or conditioning on) host, infection, and severity variables (assuming a direct effect of infection and host on outcome, as seems likely) i.e all the variables that were included in the logistic regression model in Chapter 3. To solve the problem of nonidentifiability of the models including malaria and meningitis status, I refit the models in a Bayesian framework with weakly informative priors. A student's t distribution centred on 0 with three degrees of freedom and a scale of 2.5 was used as the prior for all parameters, following Gelman et al[7]. The model was fit using the *brms* package in R[8], which acts as a front end to the Stan probabilistic programming language[9]. Four Markov-chain Monte-Carlo (MCMC) chains each with 1000 iterations and a burn-in of 500 iterations were used with default *brms* settings. Convergence was assessed using traceplots and assessing for autocorrelation using the Gelman-Rubin diagnostic ($\hat{R}$) with a target of $\hat{R} < 1.1$). Parameter estimates were expressed as medians and 95% credible intervals.

To correct for missing-data bias, missing data were imputed using multiple imputation of chained equations using default settings in the *mice* package in R[10], with each missing variable from the 18 included predicted by all other missing variables, to produce 5 imputed datasets. Models were fit using *brms* and then pooled parameter values calculated by taking medians and 95% confidence intervals of pooled posterior parameter estimates from all imputed datasets. The parameter estimates from complete-case analysis and multiple imputation are both presented.

One of the concerns of this model is that it is overfit - that is, there are so many parameters that it will fit to noise in the data rather than to the true data-generating process. To assess whether this was the case, I performed dimensionality reduction to collapse the predictor variables

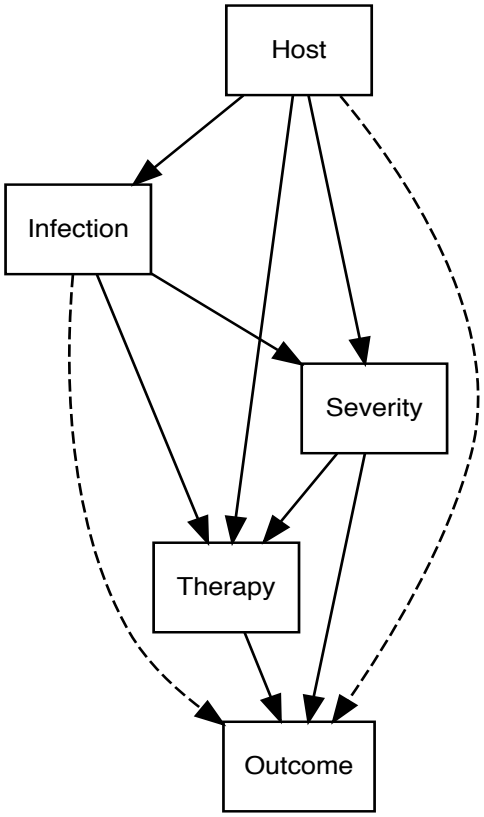

Figure 4.1: Hypothesised causal structre of death in sepsis. Host factors (e.g. age, sex, immune status) influence the type of infection; disseminated TB is more common in HIV, for example. Severity (variables quantifying e.g. shock or respiratory failure) is influenced by infection type and host factors. Therapy encodes which antimicrobials were administered and rapidity of administration of antimicrobials, and is influenced by disease severity (sicker patients may be given different therapies), host factors (HIV status may influence treatment) and the infection type (for example, malaria rapid diagnostic tests influenceing rapidity of malaria treatment). Dotted edges from host and infection to outcome are because it is not clear extita priori whether the effect of infection and host factors are entirely mediated by disease severity: in fact, even if this were the case in a theoretical sense, the available severity variables are unlikley to completely account for the causative effect of infection type on mortality and so conditioning on all available severity variables is likely to leave some residual causative effect of infection type. See text for further discussion

into a smaller number of variables, refit models using these variables, and then compared all models predictive ability using leave-one out cross validation. The dimensionality resuction technique that I used was factor analysis of mixed data (FAMD) from the *FactoMineR* package in R[11]. This technique uses principal component analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) to generate a new orthogonal coordinate system which maximises explained variance in each FAMD axis. FAMD axis one therefore explaines the most variance in the dataset, followed by FAMD axes 2 and 3, and so on. As well as reducing the dimensionality of the dataset, this technique has the advantage of ensuring an orthogonal coordinate system to tackle the problem of collinearity. The raw covariate values were used to generate these new coordianate system.

Because the exposures of interest are the therapies administered to the participants in the study, treatment variables (reciept of antibacterials, antifungals, antimalarials, antimycobacterials, and IV fluid) were left untransformed. These variables and a number of transformed FAMD variables (ranging form 1 to 5) were as predictors in new models to predict death by 28 days. The out of sample predictive ability of the models was assessed by performing leave-one-out cross validation using the *loo* package in R. This estimates the out-of sample predictive ability of the model by estimating a quantity called the expected log pointwise predictive density ($ELPD$) essentially the log of the likelihood for a new, unseen dataset conditional on the current data. This quantity is estimated using leave-one-out cross validation to produce and estimate of the $ELPD$, hereafter referred to as $ELPD_{loo}$. The standard error of $ELPD_{loo}$ for a model is also calculated and so two models can be compared by comparing the $ELPD_{loo}$ difference and standard error; if the difference is greater than twice the standard error (i.e. a 95% confidence interval, assuming normality) we can be confident that one model would be expected to have greater out-of-sample predictive ability than the other[12].

The relationship between time-to-antimicrobials and mortality was assessed, initially in bivariate associations using nonparametric locally estimates scatterplot smoothing (LOESS) regression which performs a rolling linear regression[13] and estimates the probability of death by 28 days as a function of the predictor variables. Only for antibacterials were there sufficient data to construct regression models which used time to antibacterial therapy as a predictor for death by 28 days, alongside the other treatment variables and the first three FAMD dimensions. In view of possible nonlinear relationship between time to therapy and death apparent in the bivariate plots both linear and second-order polynomial models were fit. Coefficient estimates are presented, but because interpretation of polynomial coefficients is challenging, predicted probability plots with 95% confidence intervals with the levels of the other covariates set to their mean values were plotted, using all the posterior draws to generate the median prediction and 95% confidence intervals.

Finally, to attempt to correct for confounding using a different method, a propensity-score matching approach was used to produce an unbiased estimate of the effect of receipt of TB therapy on 28-day mortality. Variables that had been identfied as being associated with mortality from the models described above, along with variables that were associated with receipt of TB therapy apparent on bivariable analysis were included in a logistic regression model to generate a propensity score. Because HIV-uninfected participants did not have a CD4 count measured, a new dichotomous variable was used which was coded as 1 for HIV-infected participants with a CD4 count below 100 cells μL^{-1} and 0 for everyone else. Participants were then matched 1:1 on this propensity score using the *MatchIt* package in R and the distribution of covariates in this new cohort examined using kernal denisty plots and histograms. Effect of TB therapy on mortality was then expressed as risk ratios, and subgroup analysis carried out to explore whether there was any effect modification of the apparent effect of TB therapy in advanced immunosupression (defined as CD4 cell count below μL^{-1}), anaemia (defined as haemoglobin below 8g dL^{-1}) or confirmed TB.

4.3 Results

Bayesian logistic regression with weakly informative priors succeding in fitting the models from Chapter 3; the inferences - particularly concerning the apaprent association between TB therapy and survival - were largely unchanged, including after muliple imputation of missing data (Table 4.1). Because of concerns about overfitting, dimensionality reduction with FA<D was carried out; the first 3 FAMD dimensions explained 34% of the variance in the dataset, a not inconsiderable amount for 18 predictor variables. The composition of FAMD dimensions one,two and three are shown in Figure 4.2A-B expressed as a plot of the squared correlation ratio (for categoriacal variables) and the squared correlation coefficient (for continuous variables) for each of the variables included in the analysis. Graphically, FAMD dimension one appeared to show an assocation with mortality (Figure 4.2C-D.

The first 5 FAMD dimensions were then used to fit models predictive of death by 28 days, along with untransformed treatment variables; the primary interest here was to see if the apparent effect of treatment administered would change under these models. Five models were fit using one, two, three, four or five FAMD dimensions; parameter estimates from these models are shown in Figure 4.3A, along with the parameter estimates form the original model using all, untransformed, parameters. Parameter estimates from treament variables were largely unchanged across, though uncertainty was markedly increased in the original models. Nevertheless, inferences were largely unchanged: we can be confident only that the odds ratio of the effect of TB treatment is different to zero.

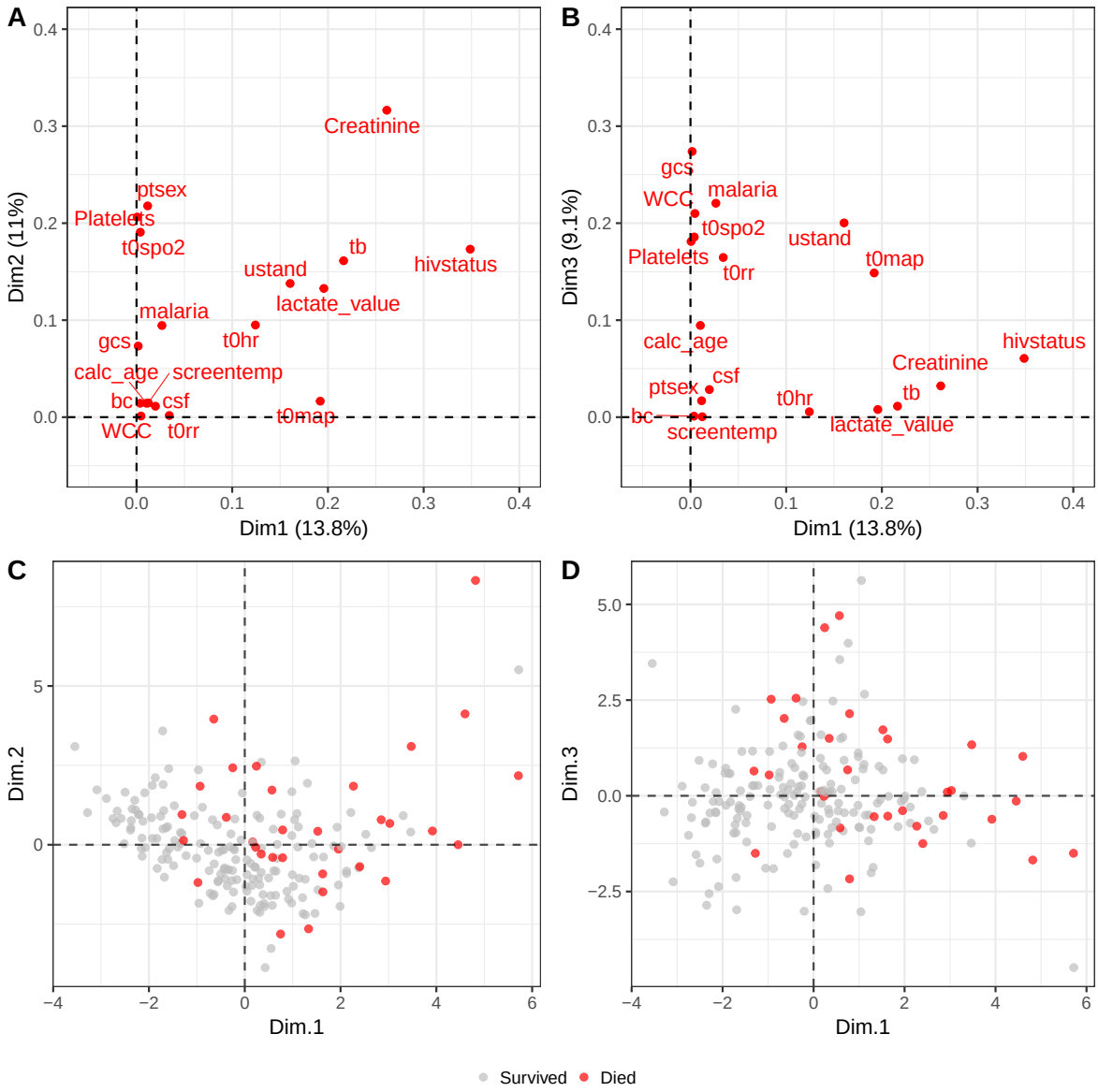

Figure 4.2: Dimensionality reduction of dataset using factor analysis of mixed data (FAMD); this is a combination of principal components analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) for categorical variables, resulting in a new orthogonal coordinate system which maximises explained variance in each FAMD axis. A and B show the squared correlation ratio (for categoriacal variables) and the squared correlation coefficient (for continuous variables) with dimensions 1 and 2 (A) or 1 and 3 (B), along wih the proportion of variance explained by each axis. C shows the location of all individuals in the FAMD space, with patients who died by 28 days coloured red to show that Dim.1 seems to be associated with mortality.

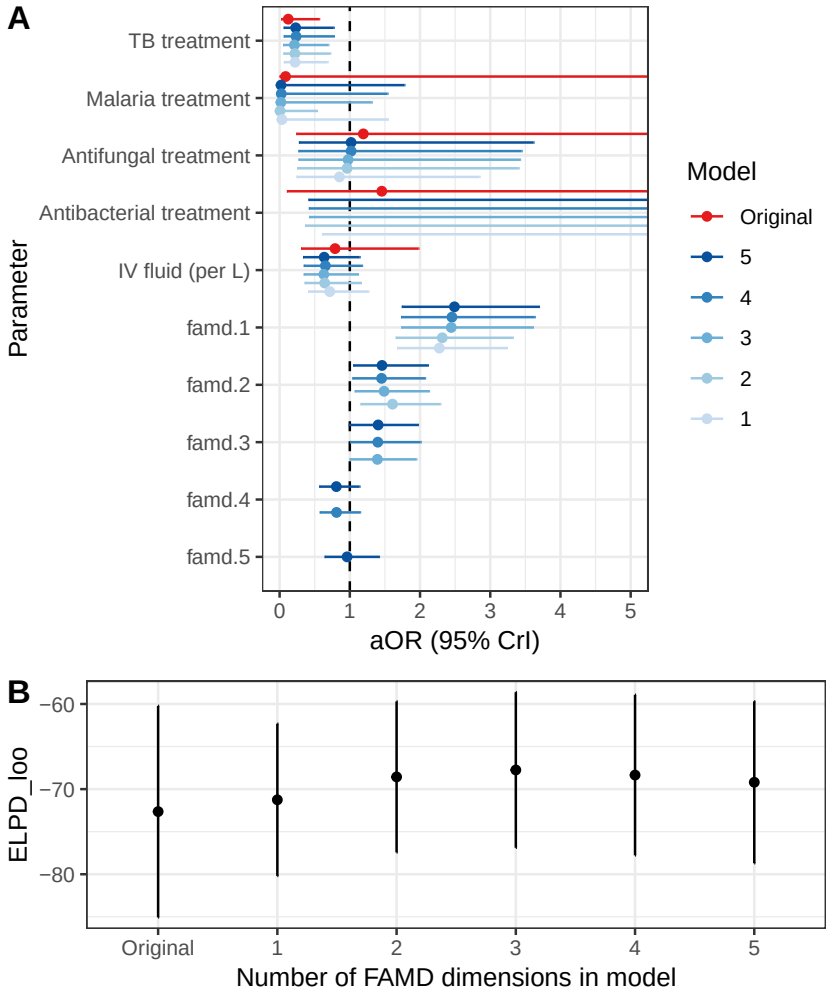

Figure 4.3: Modelling the effect of reciept of different treatments following dimensionality reduction with factor analysis of mixed data (FAMD). A (Top) shows parameter estimates for treatment variables only from the original imputed model using all raw covariate values. Models 1-5 use the first 1,2,3,4 or 5 transformed dimensions from FAMD. Parameter estimates and inferences are essentially unchanged, though there is less uncertainty generally in the estimates from the FAMD models. This would be expected as fewer parameters with less collinearity, are used. B (bottom) shows the estimated ELPD (expected log predictive denisty) from leave-one out cross validation from all the models, along with the standard error of the estimate. This is a measure of out of sample predictive accuracy: bigger (less negative) is better. One of the concerns of the original model is that it is overfit and so would have poor ELPD. In absolute terms this is true but the magnitude of the difference is much less than the standard error, meaning that out of sample prediction for all the models is broadly similar, giving confidence in the original model inferences.

The out-of-sample predictive abilit of the models was assessed using the expected log predictive density (ELPD) estimate form leave one out cross validation. In absolute terms, all FAMD models greater ELPD than the original model but any differences were small compared to the standard error of the ELPD estimate. We can not be confident that any model has different out of sample predictive accuracy and therefore can be as confident in the paraemeter etimates form th eoriginal (untransformed) model as any other.

4.3.1 Exploring time-to antibacterials and IV fluid as determinents of mortality

Exploration of bivariate associations of mortality with time to antimicrobials and volume of intravenous fluid recieved are shown in Figure 4.4, where LOESS moving linear regression provides a nonparametric estimate of probability of death by 28 days as a function of treatment variables. Time to antimalarial therapy is not shown in this plot as no patient who received antimalarial therapy died. Volume of intravenous fluid administered does has no apparent effect on 28 day mortality (Figure 4.4A). It might be expected that any effect would be most apparent in participants with shock: stratifying the analysis by shock (defined as mean arterial blood pressure below 75mmHg, Figure (Figures 4.4B) once again showed no apparent relationship. Neither time to antimycobacterial or antifungal therapy showed any apparent association though comfidence intervals are wide (Figures 4.4C and D).

There was no apparent relationshiop between time to antibacterials and2 8-day mortality up to around 40 hours, when there was a suggstion of an increased probability of death (Figure 4.4E). To explore this further, I used a logistic regression analysis, including only patients who received antibacterials (n = 207) using both linear models, fitted in a Bayesian framework as before, and, in view of a possible nonlinear effect, second order polynomial models. The estimates of the coefficients of the linear model is shown in Table 4.2 and the predicted probability of death by 28 days shown in Figure 4.4. In both cases it is not possible to fully rule in or out an effect of antibacterial delay. The 95% credible interval of the adjscted odds ratio for death per hour of antibacterial delay from the linear model crossed one (aOR 1.01 95% [CrI 0.98-1.04]) though incorporated a clinically relevant effect size, and the uncertainty in predictions from the polynomial of a late nonlinear effect of antibacterial delay are so wide that it is not possible to draw any conclusions.

4.3.2 Propensity score matching and subgroup analysis

Finally, I used propensity score matching; a different method to attempt to generate unbiased estimates of the effect of receipt of TB therapy on mortalilty. First I examined bivariate

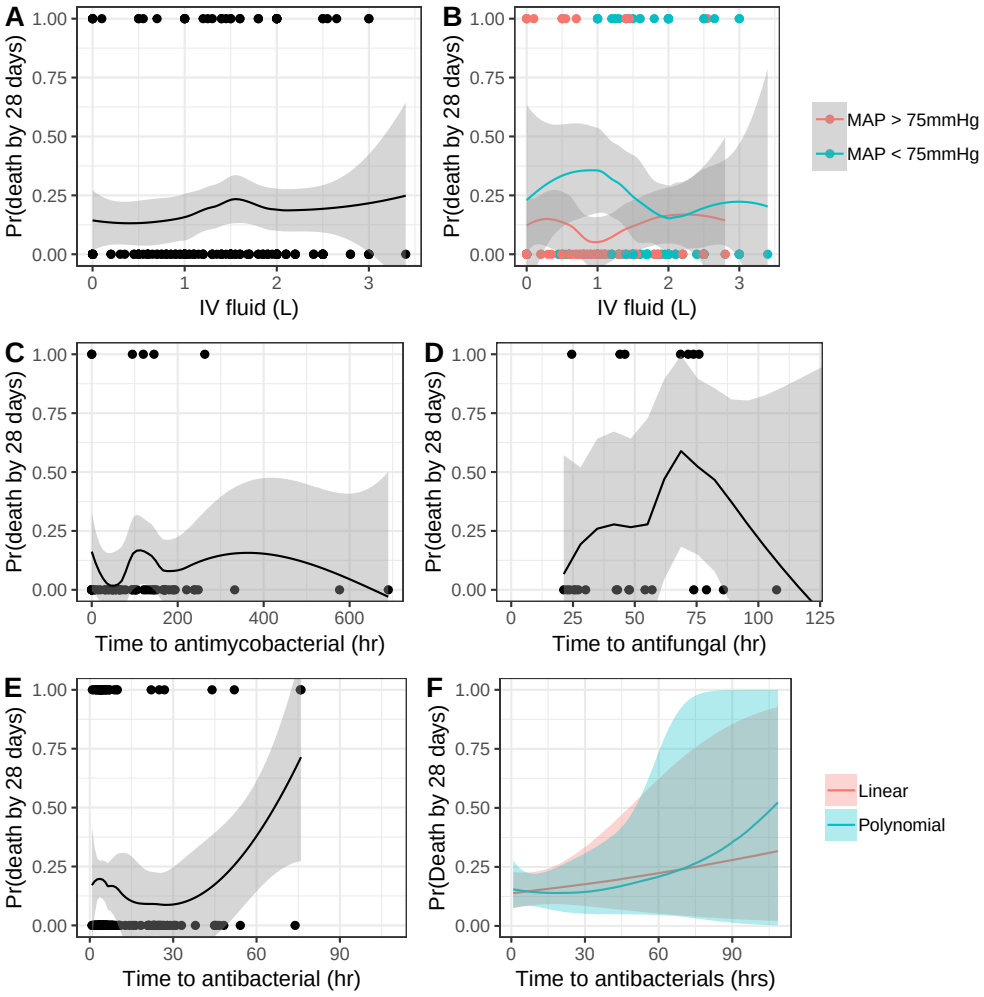

Figure 4.4: Associations of IV fliud volume and time-to-antimicrobials with death by 28 days. A-E show nonparametric regression (LOESS) of outcome (with death coded as 1 for died and 0 for survived) against various covariates; the regression line can be interpreted as the probability of death by 28 days and can be used to assess for a bivariate relationship and also the nature of any relationship (i.e. linear versus nonlinear). A: IV fluid (L), B: IV fluid stratified by presence or absence of shock (defined as MAP $<$ 75mmHg), C: Time to antimycobacterials, D: Time to antifungals E: Time to antibacterials, with a possible late, nonlinear relationship. F: Models of time-to-antimicrobials as a predictor of mortality considering time-to-antibacterials to have a linear or second order polynomial effect. In both cases the uncertainty in the effect is such that there is no convincing relationship. Overall, there is no convincing relationship between any of these variables and death by 28 days.

associations of receipt of TB therapy (Table **??** in the chapter appendix). Patients who received TB therapy were almost all HIV-infected (88% [46/52] vs 60% [95/161] in the no-TB therapy group, $p < 0.001$) with lower CD4 count (median 60 vs 123 cells μL^{-1}, $p = 0.006$) and Haemoglobin (median 9.7 vs 11.1 g dL^{-1}), and received more antimalarials (11% [6/53] vs 3% [6.172], $p = 0.037$) and IV fluids (median 1.5L vs 1.2L, p= 0.02), though most of these associations would be expected to pull an estimate of the mortality effect of TB therapy towards the null, rather than inflate an effect size. More patients with a positive diagnostic test for TB received TB therapy, as might be expected (53% [28/53] of those receiving TB therapy had a positive diagnostic test for TB, versus 28% [48/172] not receiving therapy, $p = 0.001$), though almost all the TB treatment was empiric, as the treating clinicians did not have access to urinary LAM results (which were batch processed on frozen urines) or mycobacterial blood culture results (which take up to 6 weeks to become positive).

Factors associatied with reciept of TB therapy (HIV status, CD4 count, diagnosis of TB, reciept of antimalarial therapy and volume of IV fluid received) and factors associated with mortality from the models presented above (haemoglobin, respiratory rate, oxygen saturation, inability to stand, bloodstream infection and diagnosis of malaria) were used as predictors in a logistic regression to predict reciept of TB therapy. Predictions from this model were used to generate a propensity score for each participant, and then each participant who recieved TB therapy was matched with one participant who did not to generate a new cohort, with better matching of covariates 4.6. The propeisty-socre adjucted risk ratio of survival to 28 days in this cohort upon reciept of TB therapy was 1.25 (95% CI 1.04-1.51), similar to the unadjusted estimate (Figure 4.5). Mortality benefit seemed higher in the immunosuppressed and anaemic in absolute terms, though with significant uncertainty in the estimates (4.5): RR 1.56 (95% CI 1.04-2.24) in those with haemoglobin below 8g dL^{-1} compared to 1.11 (95% CI 0.92-1.34) above 8g dL^{-1}.

4.4 Discussion

Using dimensionality reduction and Bayesian logistic regression, I present an assessment of the independent mortality effects of the treatments administered to the cohort, with an aim to inform novel antimicrobial strategies for sepsis in sSA. These approaches were used to deal with the problems of variable selection, collinearity, and nonidentifiability due to separation in logistic regression, and it possible to draw several conclusions from the results. Firstly, there is heterogeneity in outcome across diagnoses: even after controlling for disease severity, malaria was strongly associated with survival to 28 days, and meningitis with death. Mindful of the hypothesised causal structure presented above, this suggests that not all of the mortality risk

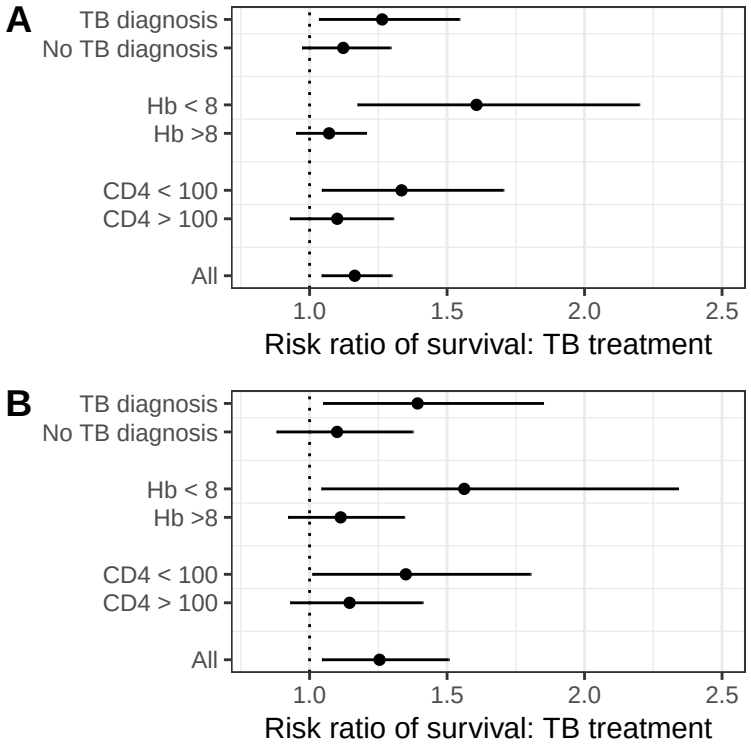

Figure 4.5: Subgroup analysis of effect of TB therapy on mortality. A (Top) shows crude (unadjusted) risk ratio for survival to 28 days; RR > 1 favours TB therapy, RR < 1 favours no TB therapy. A significant effect is seen in the immunosupressed, anaemic, and to a lesser extent, those with a confirmed diagnosis of TB. B (Bottom) shows the same analysis for the propensity-score matched cohort, showing that the overall and subgroup effects are essentially unchanged.

of death is mediated by the included disease severity markers. The reason for the low mortality of participants with malaria could be due to host factors (partial immunity) or treatment factors (rapid definitive diagnosis using point of care tests) in the context of true malaria disease, or apparent positive tests for malaria could represent incidental parasitaemia. In high-resource settings, rapid administration of antimicrobials has been shown to be associated with improved survival in sepsis[14], and all sepsis guidelines stress the importance of rapid administration of antimicrobials[15]. This is based purely on observational evidence and no RCT has ever been (or will be, given the ethical issues) carried out; these studies are all open to confounding and require adjustment for disease severity. In this study, no significant effect of time-to-antibacterials was seen, though it is important not to interpret this lack of detected effect as lack of effect. The largest study to address this question, in a high income setting (New York, USA) found an adjusted odds ratio of 1.04 (95% CI 1.02-1.05) for death per hour delay of antibiotics, and included 49,331 participants[16]. Estimates from this study are at least consistent with those, though a lack of precision here could be due to underpowering.

The most striking finding from the analysis of determinants of mortality, however, is a very strong association between receipt of TB therapy and survival. Care must be taken in interpreting this as cause and effect. Though every attempt has been made to adjust for confounding, in an observational study such as this residual unmeasured confounding is likely. It does seem, however, as though confounding would be likely to bias an estimate of the effect of TB therapy towards the null (in that clinicians might initiate TB therapy on patients who are more unwell and hence more likely to die) rather than producing a spurious protective effect. The benefit of TB therapy was not restricted to those with a confirmed diagnosis of TB, though almost all (88%) of participants who received TB therapy were HIV-infected and care should be exercised in extrapolating to the HIV-uninfected. The effect seems stronger - perhaps even limited to - those with advanced immunosuppression and/or anaemia (which itself is often associated with immunosuppression), though these conclusions are from an unadjusted analysis and should be interpreted with caution.

A protective effect of TB therapy in sepsis is plausible from prior studies: autopsy studies show that TB is under diagnosed in HIV-infected patients who die in hospital[17]. The STAMP trial[18] found a mortality benefit in some a priori subgroups of a strategy of screen-and-treat with urinary LAM for all HIV-infected inpatients, suggesting a significant burden of undiagnosed TB, and prior sepsis cohorts in sSA have found TB as a common cause of sepsis. A retrospective study of 149 HIV infected adults with sepsis in Uganda[19], 55 of whom received anti-TB therapy, found an association between receipt of TB therapy and survival in Sepsis-2 severe sepsis (hazard ratio 0.32 95% CI 0.13-0.80 from Cox proportional hazard model) but not Sepsis-2 sepsis (hazard ratio 1.24 95% CI 0.53-2.90), but is hampered by its retrospective design.

What, then, is the role of TB therapy in sepsis in sSA? The fact that the mortality benefit in this study is not restricted to those with a confirmed diagnosis of TB suggests that empiric TB therapy in sepsis or a subset of patients with sepsis (particularly those with a CD4 cell count below 50 cells μL^{-1}, or haemoglobin below 7 g dL^{-1}) could be beneficial. RCTs of empirical TB treatment have not previously been successful. The REMEMBER trial recruited outpatients with CD4 cell count below 50 cells μL^{-1} and randomised them to isoniazid preventative therapy or full TB therapy, and found no mortality benefit. STASIS found no difference in mortality between a strategy of Xpert and urine LAM screening versus empiric TB therapy in outpatients with CD4 count below 100 cells[20] μL^{-1} and TB Fast Track found no mortality benefit in empiric therapy for outpatients with CD4 cell count below 150 μL^{-1} if they were randomised to an algorithm that started TB therapy if they were assessed as high risk for TB using a combination of diagnostic tests (including urinary LAM) and clinical features (including BMI and haemoglobin)[21]. However all of these studies recruited ambulatory outpatients; it may be that inpatients have more disseminated TB, or a higher baseline risk of mortality. Empiric TB therapy for sepsis in a high-HIV/TB burden setting is a strategy that has never been assessed in an RCT.

The WHO provides guidance on empiric TB therapy in inpatients, however[22]. Hospitalised HIV-infected patients in high TB burden settings with cough and so-called "danger signs" (fever $>$ 39°C, inability to stand, respiratory rate above 30 min^{-1}, heart rate above 120 min^{-1}) should receive broad spectrum antimicrobials for 3-5 days, and, if there is no improvement, consider empiric TB therapy. This strategy was developed based largely on expert opinion, but has been shown to improve survival compared to usual care in a before-after study in South Africa[23]. Whether a 3-5 day delay will worsen outcomes in critically unwell patients with TB is unknown. There was no apparent relationship seen in this study between time to antitubercular therapy and death, but numbers were small (n= 53), and TB therapy administration was reasonably rapid, with a median of 120.6 hours from admission to administration of TB therapy; 56% (35/53) of participants received TB therapy in less than 5 days.

4.4.1 Limitations

4.5 Conclusions and further work

4.6 Appendix

Below I show bivariable associations of reciept of TB therapy and variable distributions of the propensity-score matched cohort.

Table 4.3: Bivariate associations with receipt of TB treatment

Variable	TB treatment	No TB treatment	p
Host Variables			
Age (years)	37.7 (32.5-42.9)	35.6 (26.8-43.6)	0.487
Male sex	30/53 (57%)	84/172 (49%)	0.349
HIV Infected*	**46/52 (88%)**	**97/161 (60%)**	**<0.001**
Taking ART†	35/46 (76%)	82/97 (85%)	0.250
CD4 count† (μL^{-1})	**60.0 (26.2-114.8)**	**123.0 (39.0-274.0)**	**0.006**
Haemoglobin (x10^9 g dL^{-1})	**9.7 (7.4-11.3)**	**11.1 (8.6-13.9)**	**0.001**
Severity Variables			
Temperature (°C)	38.5 (38.0-39.2)	38.4 (37.9-39.0)	0.487
Heart rate (min^{-1}))	125.0 (110.0-134.0)	119.5 (99.8-132.0)	0.051
Systolic BP (mmHg)	92.0 (81.0-107.0)	99.0 (86.0-120.0)	0.133
Diastolic BP (mmHg)	67.0 (56.0-71.0)	65.0 (57.0-78.8)	0.486
Mean arterial BP (mmHg)	76.0 (65.3-83.7)	77.2 (65.1-91.2)	0.272
Respiratory rate (min^{-1}))	34.0 (30.0-38.0)	34.0 (32.0-37.0)	0.503
Oxygen saturation (%)	96.0 (94.0-98.0)	96.0 (95.0-98.0)	0.871
GCS	15.0 (15.0-15.0)	15.0 (15.0-15.0)	0.566
Unable to stand	13/53 (25%)	50/172 (29%)	0.601
Lactate (mmol L^{-1})	3.2 (2.4-4.9)	3.4 (2.2-5.3)	0.796
White cell count (x10^9 L^{-1})	6.4 (4.6-9.1)	6.6 (4.3-11.7)	0.595
Platelet count (x10^9 L^{-1})	225.5 (146.8-303.2)	215.0 (145.0-296.0)	0.498
Bicarbonate (mmol L^{-1})	18.0 (16.0-21.0)	20.0 (17.0-22.5)	0.065
Urea (mmol L^{-1})	5.0 (3.8-8.7)	4.6 (3.3-7.7)	0.174
Creatinine (mmol L^{-1})	76.0 (59.0-105.0)	75.5 (59.0-102.2)	0.824
Diagnosis			
BSI	4/53 (8%)	20/172 (12%)	0.611
TB	**28/53 (53%)**	**48/172 (28%)**	**0.001**
Malaria	6/53 (11%)	15/172 (9%)	0.592
Meningitis	1/53 (2%)	3/172 (2%)	1.000
No diagnosis	**15/53 (28%)**	**96/172 (56%)**	**0.001**
Treatment Received			
Antibacterials	47/53 (89%)	160/172 (93%)	0.383
Time to Antibacterials (hr)	5.1 (3.8-9.7)	5.4 (3.6-13.4)	0.844
Antifungals	8/53 (15%)	18/172 (10%)	0.337
Time to Antifungals (hr)	45.4 (25.2-60.6)	50.9 (33.2-78.3)	0.243
Antimalarials	**6/53 (11%)**	**6/172 (3%)**	**0.037**
Time to Antimalarials (hr)	4.5 (3.0-11.7)	12.5 (3.3-21.7)	0.631
IV fluid (ml)	**1.5 (1.0-2.0)**	**1.2 (0.5-2.0)**	**0.020**

Note:
BP = Blood pressure, GCS = Glasgow coma scale. Numeric variables are presented as median (IQR) and categorical variables as proportions. P-values are from Kruskal-Wallace test for continuous variables and Fisher's exact test for categorical variables.

* Participants with HIV status unknown not included in this row

† Includes only HIV-infected participants

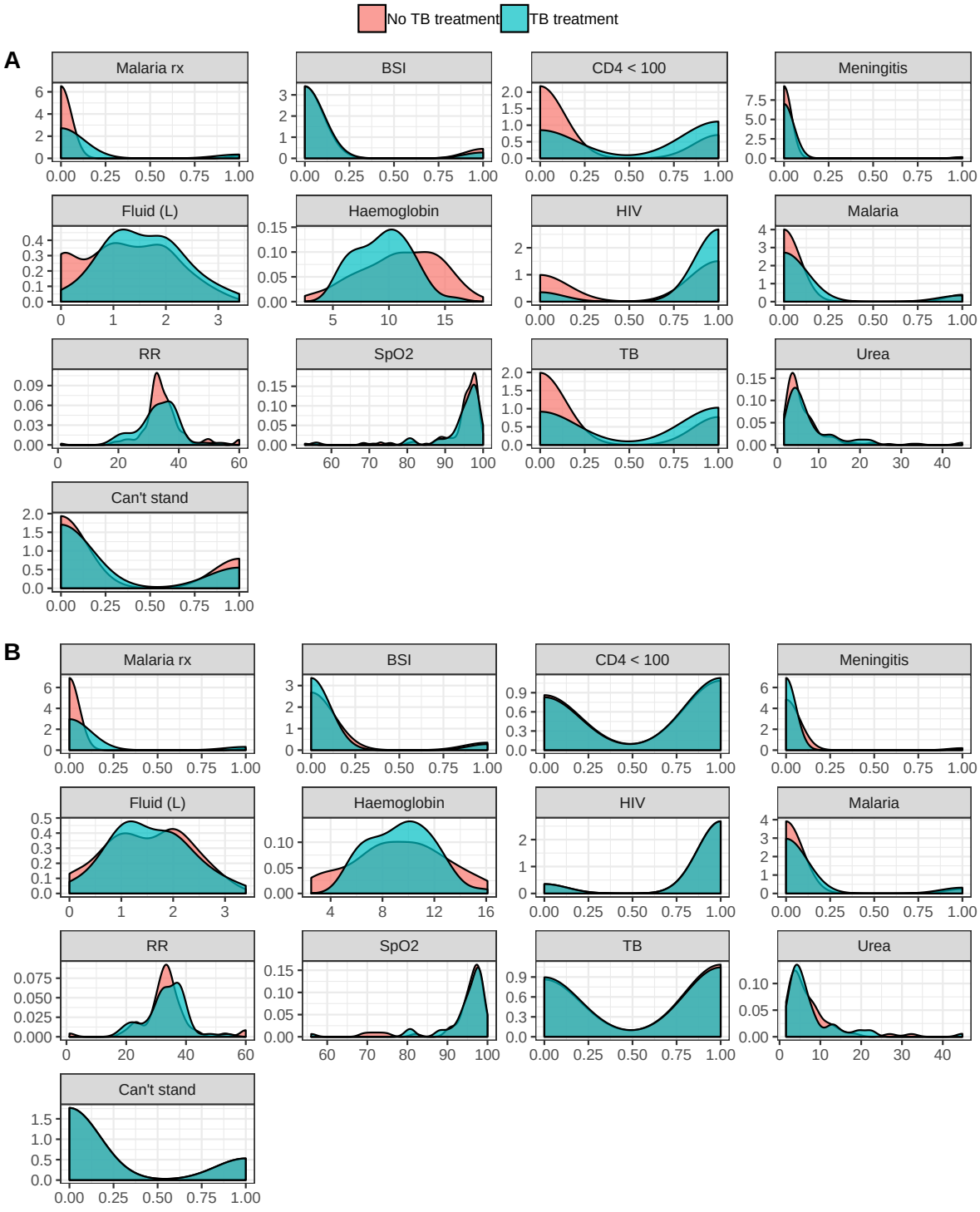

Figure 4.6: Variable distributins following propensity score matching. A: original cohort. B: Propesnity score matched cohort. BSI = bloodstream infection, RR = respiratory rate, Sp02 = Capilliary oxygen saturation, TB = tuberculosis. Categrorical variables (Malaria rx, BSI, CD4 < 100, Meningitis, HIV, Malaria, TB, Can't stand) are coded as 1 for present and 0 for absent.

Table 4.1: Unadjusted and adjusted odds ratios of death by 28 days

	aOR (95% CrI	
Variable	CCA	Imputed
Host Variables		
Age (per 5 years increase)	0.92 (0.66-1.25)	0.87 (0.65-1.14)
Male sex (vs female)	0.91 (0.22-3.51)	0.67 (0.18-2.36)
HIV Infected (vs uninfected)	0.21 (0.03-1.22)	0.32 (0.06-1.59)
Haemoglobin (per g dL^{-1})	**0.71 (0.54-0.91)**	**0.69 (0.52-0.90)**
Severity Variables		
Temperature (per °C)	0.72 (0.30-1.73)	0.56 (0.26-1.18)
Heart rate (per 10 min^{-1})	1.20 (0.87-1.72)	1.14 (0.85-1.57)
Mean arterial BP (per 10 mmHg)	1.20 (0.73-1.93)	1.14 (0.70-1.80)
Respiratory rate (per 10 min^{-1})	**0.25 (0.08-0.66)**	**0.38 (0.16-0.88)**
Oxygen saturation (per 5%)	0.73 (0.48-1.11)	**0.67 (0.45-0.99)**
GCS (per 1 unit)	0.76 (0.50-1.12)	0.75 (0.51-1.10)
Unable to stand	**13.79 (2.88-74.50)**	**13.64 (3.35-64.82)**
Lactate (per 1 mmol L^{-1})	1.12 (0.91-1.39)	1.13 (0.92-1.41)
White cell count (per $1x10^9$ L^{-1})	0.96 (0.84-1.07)	0.94 (0.83-1.05)
Platelet count (per $100x10^9$ L^{-1})	1.13 (0.70-1.78)	0.94 (0.60-1.43)
Bicarbonate (per 1 mmol L^{-1})	0.97 (0.81-1.17)	0.95 (0.81-1.11)
Urea (per 1 mmol L^{-1})	1.20 (1.00-1.45)	1.17 (1.00-1.37)
Creatinine (per 10 mmol L^{-1})	0.99 (0.92-1.08)	0.99 (0.93-1.08)
Diagnosis		
BSI (vs no BSI)	**0.04 (0.00-0.48)**	**0.04 (0.00-0.40)**
TB (vs no TB)	1.12 (0.25-5.00)	0.72 (0.18-2.69)
Malaria (vs no malaria)	0.01 (0.00-2.27)	**0.00 (0.00-0.41)**
Meningitis (vs no meningitis)	**68.53 (1.29-27384.82)**	**37.00 (1.03-6237.92)**
Treatment Received		
Received antibacterial (vs none)	8.38 (0.20-6631.38)	1.46 (0.10-30.00)
Received antifungal (vs none)	1.39 (0.23-8.60)	1.19 (0.24-5.65)
Received antimalarial (vs none)	0.03 (0.00-8.68)	0.08 (0.00-13.22)
Received antimycobacterial (vs none)	**0.11 (0.02-0.58)**	**0.12 (0.02-0.56)**
IV fluid (per L)	0.82 (0.29-2.21)	0.79 (0.31-1.98)

Note:
BP = Blood pressure, GCS = Glasgow coma scale, BSI = Bloodstream infection, TB = tuberculosis. All odds ratios are for as increase in the variables shown.

Table 4.2: Adjusted odds ratio of death by 28 days per hour delay in antibacterials

Variable	aOR (95% CrI)
Time to antibacterials (per hour)	1.01 (0.98-1.04)
IV fluid (per L)	0.65 (0.36-1.16)
Received antimalarial (vs none)	0.02 (0.00-1.37)
Received antifungal (vs none)	1.03 (0.28-3.49)
Received antimycobacterial (vs none)	0.18 (0.05-0.61)
famd.1	2.56 (1.81-3.73)
famd.2	1.43 (1.04-2.01)
famd.3	1.38 (1.00-1.90)

Note:
The variables famd1,2 and 3 are the three transformed dimensions following dimensionality reduction using factor analysis of mixed data that account for the most variability in the dataset.

Chapter 5

ESBL-E carriage in Malawian adults in health and disease

Placeholder

5.1 Chapter Overview

5.2 Introduction and chapter aims

5.3 Methods

5.4 Results

5.4.1 Study population

5.4.2 Exposures during the study period

5.4.3 ESBL-E colonisation

5.4.4 Associations of ESBL colonisation

5.5 Discussion

5.5.1 Limitations

5.6 Conclusions and further work

Chapter 6

Whole genome sequencing of ESBL *E. coli* carriage isolates

Placeholder

6.1 Chapter overview

6.2 Methods

6.2.1 Bioinformatic pipeline

6.2.2 Global *E. coli* collection

6.2.3 Statistical analysis

6.3 Results

6.3.1 Samples and quality control

6.3.2 Phylogroup, MLST and core genome phylogeny of study isolates

6.3.3 Study isolates in a global context

6.3.4 Antimicrobial resistance determinants

6.3.4.1 β-lactam resistance

6.3.4.2 Quinolone resistance

6.3.4.3 Aminoglycoside resistance

6.3.4.4 Chloramphenicol, co-trimoxazole, tetracycline and other resistance determinants

6.3.4.5 Clustering and lineage association of AMR determinants

6.3.5 Plasmid replicons

6.3.6 Testing metadata associations: SNP distance, hierBAPS sequence clusters and ESBL-clusters

6.3.6.1 Hierarchical BAPS clustering of core gene pseudosequences

6.3.6.2 ESBL-clusters

6.3.6.3 Assessing for healthcare-associated lineages

6.3.6.4 Assessing for within-patient conservation of lineage or MGE

6.4 Discussion

Chapter 7

Genomics I

Chapter 8

Longitudinal models of ESBL-E carriage

Placeholder

8.1 Chapter Overview

8.2 Introduction and chapter aims

8.3 Methods

8.3.1 Developing the models used in this chapter

8.3.2 General form of likelihood

8.3.3 Markov model likelihood

8.3.4 Incorporating covariates: a proportional hazard model

8.3.5 Building and fitting models

8.3.6 Assessing goodness of fit

8.3.7 Exploring differences in carriage dynamics by bacterial species and *E. coli* genotype

8.3.8 Simulations from the posterior

8.4 Results

8.4.1 The effect of antibacterials and hospitalisation on ESBL-E carriage

8.4.2 Exploring bacterial species and genotype differences in carriage dynamics

8.4.3 Simulation of different antibacterial and hospitalisation scenarios

8.5 Discussion

8.5.1 Limitations

8.6 Conclusion and further work

8.7 Appendix

References

1 Harrell FE. *Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis*. Springer 2001.

2 Tibshirani R. Regression Shrinkage and Selection via the Lasso. 1996;**58**:267–88. doi:10.2307/2346178

3 Hastie T, Tibshirani R, Friedman JH(H. *The elements of statistical learning : data mining, inference, and prediction*. Springer 2009.

4 Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology* 2013;**177**:292–8. doi:10.1093/aje/kws412

5 Pearl J, Glymour M, Jewell NP. *Causal inference in statistics : a primer*. John Wiley & Sons Ltd 2016.

6 Textor J, Zander B van der, Gilthorpe MS *et al.* Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* 2017;**45**:dyw341. doi:10.1093/ije/dyw341

7 Gelman A, Carlin JB, Stern HS *et al. Bayesian data analysis*. 3rd ed. Chapman; Hall/CRC 2004.

8 Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 2017;**80**:1–28. doi:10.18637/jss.v080.i01

9 Carpenter B, Gelman A, Hoffman MD *et al.* Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 2017;**76**:1–32. doi:10.18637/jss.v076.i01

10 Buuren S van, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011;**45**:1–67. doi:10.18637/jss.v045.i03

11 Lê S, Josse J, Husson F. FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical Software* 2008;**25**:1–18. doi:10.18637/jss.v025.i01

12 Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 2017;**27**:1413–32. doi:10.1007/s11222-016-9696-4

13 Cleveland WS, Devlin SJ. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* 1988;**83**:596–610. doi:10.1080/01621459.1988.10478639

14 Kumar A, Roberts D, Wood KE *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical care medicine* 2006;**34**:1589–96. doi:10.1097/01.CCM.0000217961.75225.E9

15 Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Medicine* 2018;**44**:925–8. doi:10.1007/s00134-018-5085-0

16 Seymour CW, Gesten F, Prescott HC *et al.* Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine* 2017;**376**:2235–44. doi:10.1056/NEJMoa1703058

17 Gupta RK, Lucas SB, Fielding KL *et al.* Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings. *AIDS* 2015;**29**:1987–2002. doi:10.1097/QAD.0000000000000802

18 Gupta-Wright A, Corbett EL, Oosterhout JJ van *et al.* Rapid urine-based screening for tuberculosis in HIV-positive patients admitted to hospital in Africa (STAMP): a pragmatic, multicentre, parallel-group, double-blind, randomised controlled trial. *Lancet (London, England)* 2018;**392**:292–301. doi:10.1016/S0140-6736(18)31267-4

19 Hazard RH, Kagina P, Kitayimbwa R *et al.* Effect of Empiric Anti–Mycobacterium tuberculosis Therapy on Survival Among Human Immunodeficiency Virus–Infected Adults Admitted With Sepsis to a Regional Referral Hospital in Uganda. *Open Forum Infectious Diseases* 2019;**6**. doi:10.1093/ofid/ofz140

20 Blanc F-X, Badje2 AD, Bonnet M *et al.* Systematic vs test-guided tuberculosis treatment: data of the STASIS randomized trial. In: *Conference on retroviruses and opportunistic infections*. 2018.

21 Grant A, Charalambous S, Tlali M *et al.* Empirical TB Treatment in Advanced HIV Disease: Results of the TB Fast Track Trial. In: *Conference on retroviruses and opportunistic infections*. 2016.

22 World Health Organisation. Improving the diagnosis and treatment of smear-negative pulmonary and extrapulmonary tuberculosis among adults and adolescents: recommendations

for HIV-prevalent and resource-constrained settings. 2007.

23 Holtz TH, Kabera G, Mthiyane T *et al.* Use of a WHO-recommended algorithm to reduce mortality in seriously ill patients with HIV infection and smear-negative pulmonary tuberculosis in South Africa: an observational cohort study. *The Lancet Infectious diseases* 2011;**11**:533–40. doi:10.1016/S1473-3099(11)70057-3