

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

Preface	9
1 Introduction	11
1.1 Chapter Overview	13
1.2 Sepsis in sub-Saharan Africa	13
1.3 ESBL-E in sub-Saharan Africa	13
1.4 Conclusions	13
1.5 Thesis overview	13
1.6 Appendix	13
1.7 References	13
2 Methods	15
2.1 Chapter Overview	17
2.2 Study site	17
2.3 Clinical Study	17
2.4 Diagnostic Laboratory Procedures	17
2.5 Molecular methods	17
2.6 Bioinformatics	17
2.7 Statistical Analysis	17
2.8 Study Team	17
2.9 Data Collection and Storage	17
2.10 Ethical Approval, Consent and Participant Remuneration	17
3 A clinical and microbiological description of sepsis in Blantyre, Malawi	19
3.1 Chapter overview	20
3.2 Introduction and chapter aims	20
3.3 Methods	20
3.4 Results	20
3.5 Discussion	20

3.6	Conclusions and further work	20
4	Exploratory modelling of sepsis outcome	21
4.1	INtro	21
4.2	Methods	23
4.3	Results	26
4.4	Discussion	34
5	ESBL-E carriage in Malawian adults in health and disease	39
5.1	Chapter Overview	40
5.2	Introduction and chapter aims	40
5.3	Methods	40
5.4	Results	40
5.5	Discussion	40
5.6	Conclusions and further work	40
6	Whole genome sequencing of ESBL <i>E. coli</i> carriage isolates	41
6.1	Chapter overview	43
6.2	Methods	43
6.3	Results	43
6.4	Discussion	43
6.5	Appendix	43
7	Genomics I	45
8	Longitudinal models of ESBL-E carriage	47
8.1	Chapter Overview	49
8.2	Introduction and chapter aims	49
8.3	Methods	49
8.4	Results	49
8.5	Discussion	49
8.6	Conclusion and further work	49
8.7	Appendix	49
	References	51

List of Tables

4.1	Unadjusted and adjusted odds ratios of death by 28 days	28
4.2	Adjusted odds ratio of death by 28 days per hour delay in antibacterials . . .	30
4.3	Bivariate associations with receipt of TB treatment	37

List of Figures

4.1	Hypothesised causal structure of mortality in sepsis	24
4.2	Dimensionality reduction using FAMD	31
4.3	Modelling effects of treatments following dimensionality reduction.	32
4.4	Associations of time to antimicrobials and death by 28 days	33
4.5	Subgroup analysis of effect of TB therapy on mortality.	36

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-Saharan 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 identification 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

2.4.2.4 Sputum Xpert

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 INtro

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). These variables (18 in total) have been selected *a priori* and are shown in Table ?? in the results section. I chose to use the method of factor analysis of mixed data (FAMD) from the *FactoMineR* package in R[7] to perform dimensionality reduction on these variables. 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. 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.

Non-normally distributed variables (temperature, white cell count, urea, creatinine, lactate) were identified as such by examination of kernel density plots and transformed with a natural logarithm prior to FAMD, and all variables mean centered and scaled using standard deviation; two variables (oxygen saturation and Glasgow coma score [GCS]) remained nonnormal following logarithmic transformation and were recoded as binary variables with a cut off of less than 92% (for oxygen saturation) and less than 15 (for GCS). Because CD4 count was unavailable for HIV-uninfected participants, CD4 count was recoded as a binary dummy variable to be equal to 1 for HIV-infected participants with a CD4 cell count less than 100 cells μL^{-1} and 0 for all other participants. The first three FAMD dimensions were used as predictors of mortality in a logistic regression model, which also included diagnosis (TB, BSI, meningitis, malaria) and treatment (whether the participant was administered antibacterials, antimycobacterial therapy, antifungals or antimalarials). Diagnostic variables were not transformed as they are largely orthogonal (i.e. it is not possible to reduce their dimensionality to any great extent

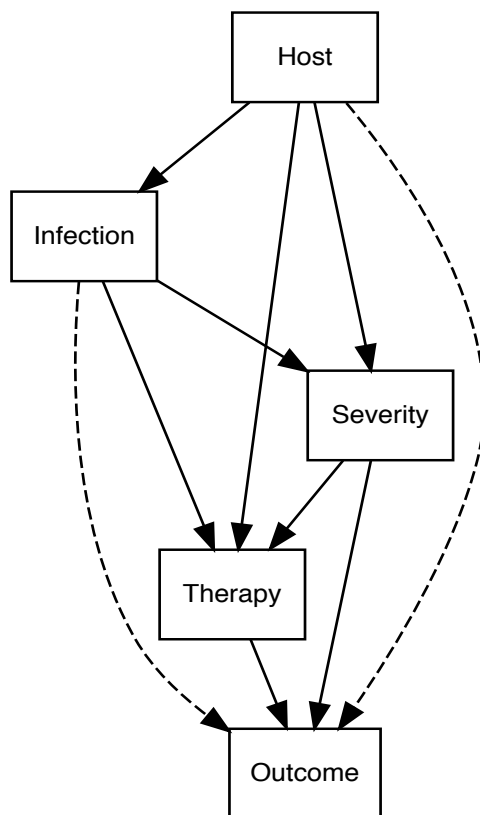


Figure 4.1: Hypothesised causal structure of death in sepsis. Host factors (e.g. age, sex, immune status) influence the type of infection; disseminated TB is more common in HIV, for example. Severity (variables quantifying e.g. shock or respiratory failure) is influenced by infection type and host factors. 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 influencing rapidity of malaria treatment). Dotted edges from host and infection to outcome are because it is not clear *a 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 unlikely 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

without losing information). Volume of intravenous fluid administered over 6 hours was included as a linear continuous predictor in this model. In all cases receipt of antimicrobials refers to antimicrobials received whilst in hospital; participants ostensibly on TB treatment prior to admission, for example, were coded as no TB treatment in the analysis.

Because of nonidentifiability of the model under a maximum likelihood framework, Bayesian regression using the *brms* package in R[8] (a front end to the Stan probabilistic programming language[9]) was used. No patients with malaria died, which meant that it is not possible for standard (maximum likelihood) models to estimate parameters for the effect of malaria mortality. Excluding this variable could result in biased estimates, but it is possible to fit the model in a Bayesian framework by specifying weakly informative priors on the parameters. In the broadest sense, we use our knowledge (that adults do die of malaria) to set priors that pull the parameter estimates to an identifiable value. Student's t distribution centred on 0 with three degrees of freedom and a scale of 2.5 were used, following Gelman et al[10]. 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 calculated from the posterior using medians and 95% confidence intervals.

Missing data were imputed using multiple imputation of chained equations using default settings in the *mice* package in R[11], with each missing variable from the 18 included predicted by all other missing variables, to produce 10 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 association between association of reduced mortality and receipt of TB therapy was explored in crude (unadjusted) subgroup analysis, as the dataset was not large enough to explore this association by fitting interaction terms. The magnitude of effect modification of TB therapy was expressed as risk ratios in several subgroups: those with and without a confirmed diagnosis of TB, and at varying cut offs of CD4 count and haemoglobin to quantify immunosuppression.

Finally, 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[12] 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 (mean-centred and scaled as before) as a predictor for death by 28 days, alongside 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,

as before. 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.

4.3 Results

TB: This is clearly a finding that is very open to confounding so the associations of receipt of TB treatment were explored (Table 4.3 in the appendix to this chapter); 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).

To explore the possibility that the apparent association of TB therapy with survival is driven by an effect in those with a diagnosis of TB only, crude (unadjusted) subgroup analysis was carried out and effect mortality estimates stratified by both TB diagnosis and receipt of anti-TB therapy are shown in Figure 4.5; the association of TB therapy with improved survival was similar though slightly higher in the confirmed TB group (risk ratio, RR, for survival 1.26 [95% CI 1.03 - 1.54] in confirmed TB versus 1.12 [95% CI 0.97-1.29]). In contrast, mortality benefit of TB therapy did seem to be more pronounced in advanced immunosuppression, with significantly larger effect in those with a CD4 count of below 50 cells μL^{-1} (RR 1.62 [95% CI 1.11-2.37]) compared to above 50 (1.07 [95% CI 0.93-1.23]) and with haemoglobin below 7 g dL^{-1} (RR 1.61 [95% CI 1.17-2.20]) compared to above (RR 1.07 [95% CI 0.951-1.21])

Kruskal-Wallis test of time to treatment with all antimicrobials found no association with 28-day mortality, and no association with volume of intravenous fluid administered over 6 hours. Further exploration of bivariate associations of mortality with these continuous treatment variables 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. No relationship is apparent, thanks to wide confidence intervals except possibly for antibacterial therapy (Figures 4.4D and E): any effect would seem to be only apparent on delay on antimicrobial administration beyond around 40 hours post-admission, with likely non-linearity in effect size as a function of antimicrobial delay. Similarly, volume of intravenous fluid administered does have no apparent effect on 28 day mortality (Figure 4.4B). 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 4.4C) once again showed no apparent relationship.

To explore these associations further, I used a logistic regression analysis, with a primary aim of describing the effect of treatments administered (antimicrobials and fluid) on mortality though as described above, there were several challenges. The problems of collinearity and variable selection were addressed with dimensionality reduction using factor analysis of mixed data (FAMD) on host and severity variables. First, temperature, white cell count, urea, creatinine, lactate were transformed with natural logarithms as their distribution was non-normal on inspection of histograms and kernel density plots. The distributions of oxygen saturation and GCS were very non-normal so were dichotomised into two categories each: GCS as either 15 or less than 15, and oxygen saturation as either above 92% or equal to or below 92%. The composition of first 3 FAMD dimensions in terms of squared correlation ratio (for categorical variables) and the squared correlation coefficient (for continuous variables) are shown in Figure ?? and explained 39% of the variance in the dataset of 18 variables. Furthermore, the dimension provides some discrimination in terms of mortality (Figure 4.2C) and dimensions one and two provide some discrimination in terms of diagnosis, particularly between TB and malaria (Figure 4.2D and E).

The first three FAMD dimensions, along with diagnosis and treatment variables were used as explanatory variables in a logistic regression model to predict mortality, though the interest was primarily to correct effect size estimates for confounding rather than predicting outcome. Dimensionality reduction was not undertaken on diagnosis variables as they are largely orthogonal, and also to maintain interpretability. Because no patients with malaria died, the standard maximum likelihood estimation of a logistic regression model failed, so Bayesian logistic regression with weakly informative priors was used following imputation of missing data to form 10 imputed datasets, as described in methods, above. MCMC diagnostics showed good sampling of the posterior: \hat{R} was less than 1.1, traceplots showed good mixing of chains (Figure ?? in the chapter appendix) and there were no divergences of the sampler. Parameter estimates and 95% credible intervals from this model are shown in Table ??, and conclusions from univariate associations are largely unchanged: we can be confident that malaria is associated with survival, meningitis with death (though with very wide credible intervals reflecting the small number of cases), and administration of TB

therapy with survival, following adjustment for the included confounders. The dataset is too small to effectively model interaction terms that would be the equivalent of the bivariate subgroup analyses presented above to test the hypothesis that fluid administration would have a different effect in shock, and that TB therapy would have a different effect in confirmed TB or immunosuppression or anaemia; we will have to be content with the unadjusted analysis. Propensity score analysis (see further work) may be able to produce adjusted estimates from subgroup analysis/

I then went on to model the effect of antibacterial delay, including only patients who received antibacterials ($n = 207$) using both linear models, and, in view of a possible nonlinear effect, second order polynomial models, both in complete case analysis and following imputation of missing data as before. These data were not included in the full model as it would mean discarding the participants who received no antibacterials. The estimates of the coefficients of these models are shown in Table ?? and the predicted probability of death by 28 days shown in Figure ?. We have low confidence that estimates of effect size of antimicrobial delay were different from 1 from simple linear models and, though interpreting the coefficients of linear models is difficult, the confidence intervals for the polynomial models are so wide that it is not possible to safely rule in or out a late effect of antimicrobial delay.

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

Variable	aOR (95% CrI)	
	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)

Table 4.1: Unadjusted and adjusted odds ratios of death by 28 days (*continued*)

Variable	aOR (95% CrI	
	CCA	Imputed
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.

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

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.

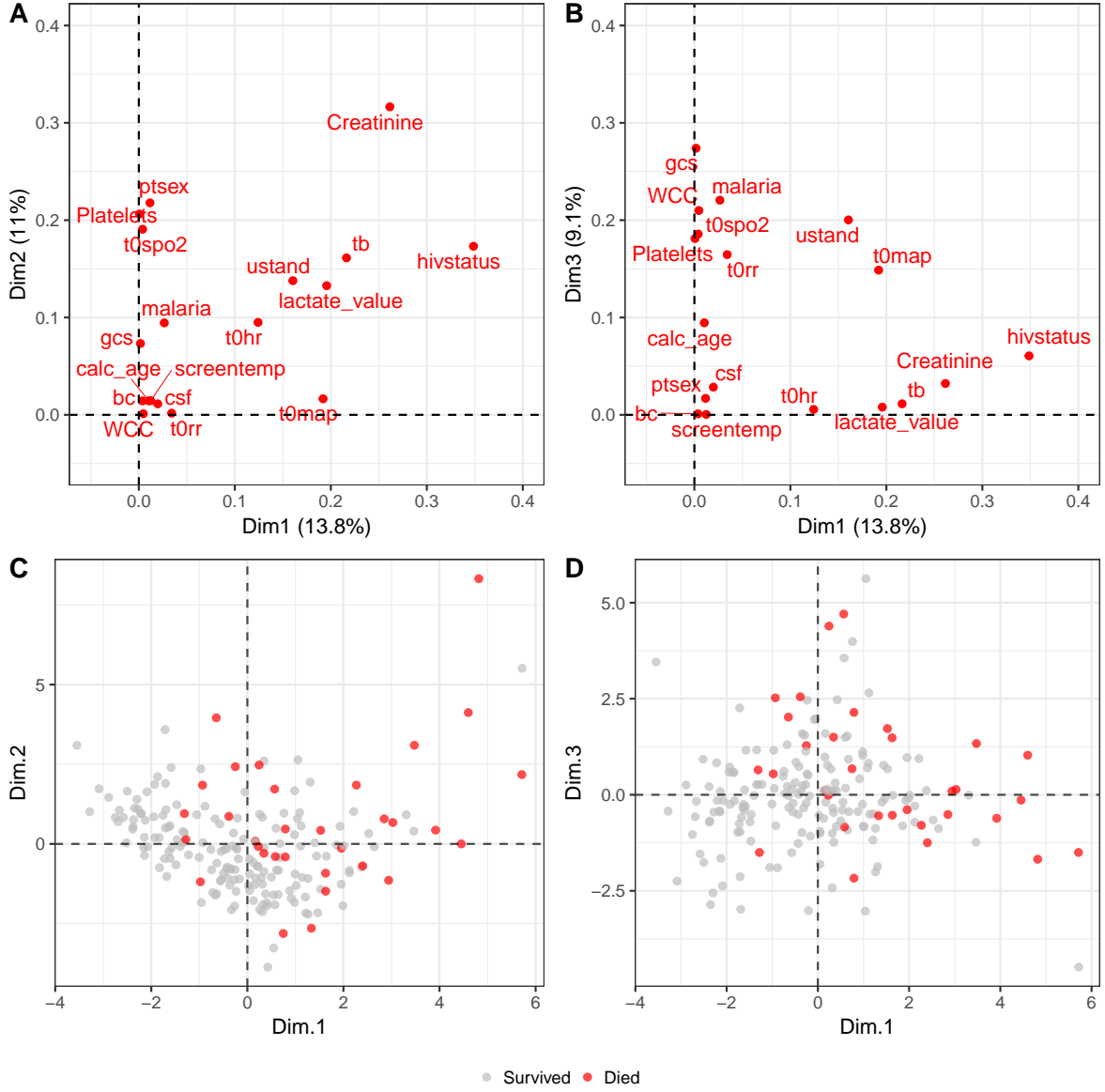


Figure 4.2: Dimensionality reduction of dataset using factor analysis of mixed data (FAMD); this is a combination of principal components analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) for categorical variables, resulting in a new orthogonal coordinate system which maximises explained variance in each FAMD axis. A and B show the squared correlation ratio (for categorical variables) and the squared correlation coefficient (for continuous variables) with dimensions 1 and 2 (A) or 1 and 3 (B), along with 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.

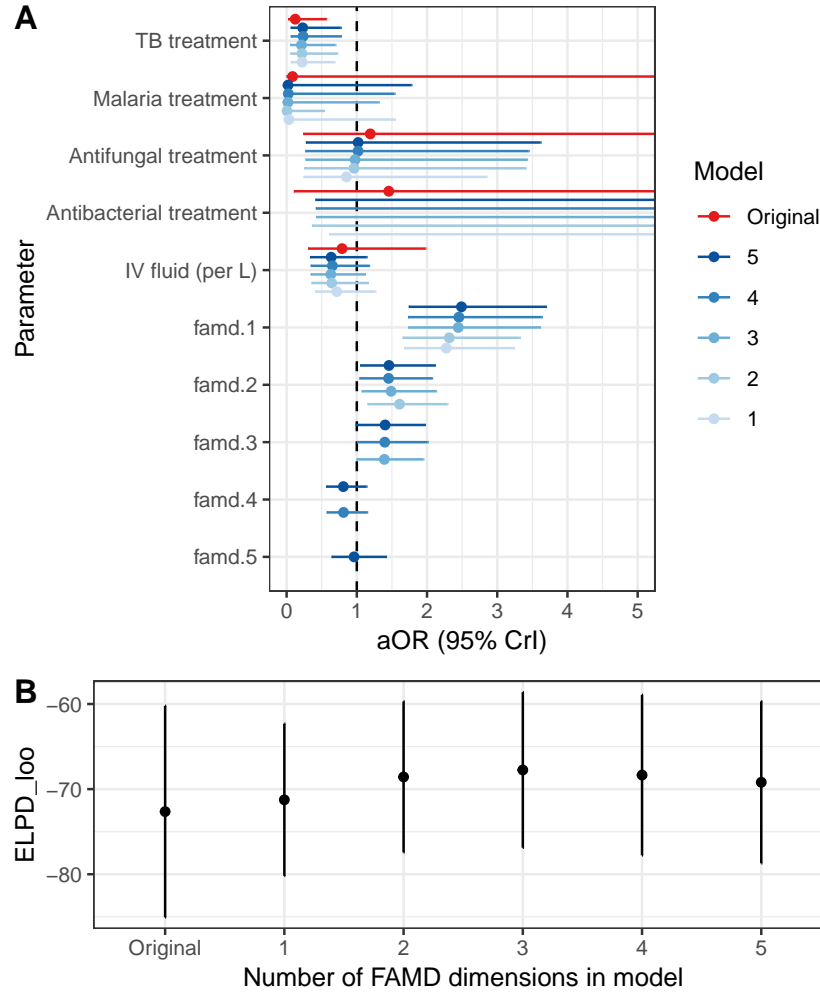


Figure 4.3: Modelling the effect of receipt of different treatments following dimensionality reduction with factor analysis of mixed data (FAMD). A (Top) shows parameter estimates for treatment variables only from the original imputed model using all raw covariate values. Models 1-5 use the first 1,2,3,4 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 density) 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. 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.

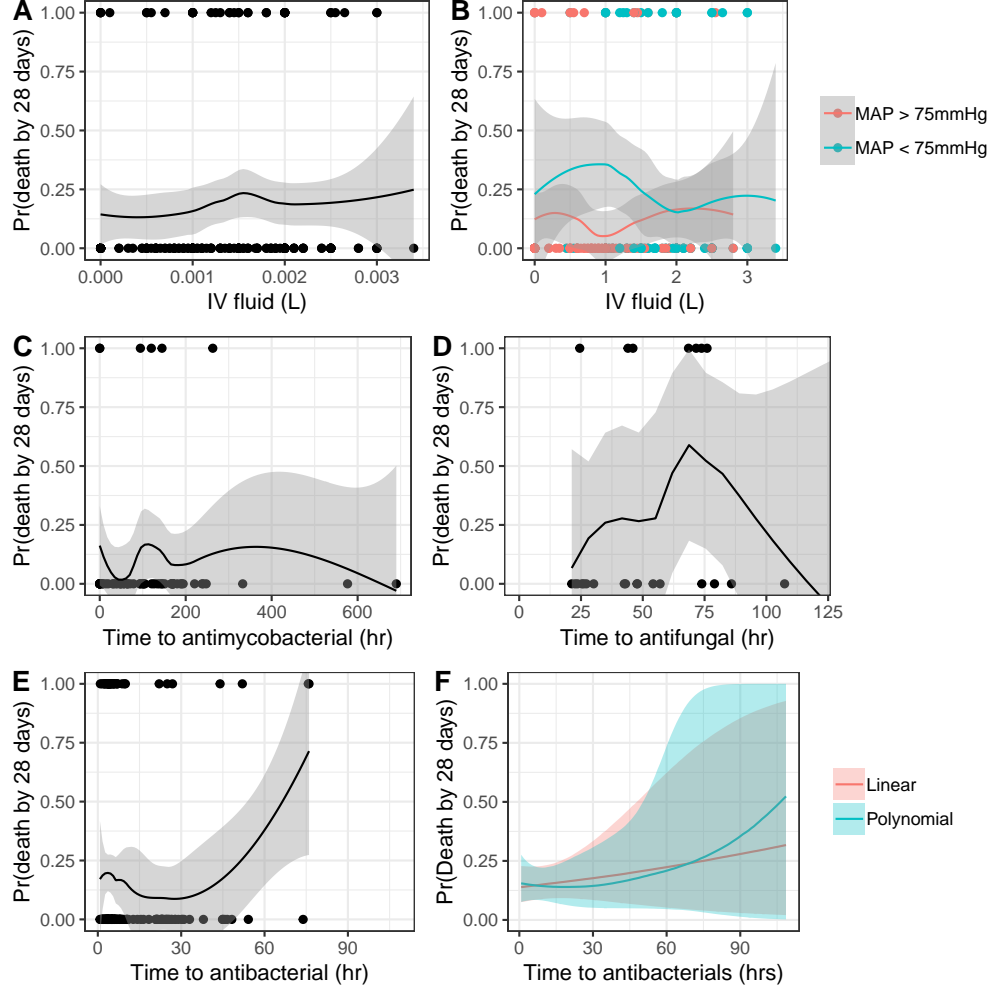


Figure 4.4: Associations of IV fluid volume and time-to-antimicrobials with death by 28 days. A-E show nonparametric regression (LOESS) of outcome (with death coded as 1 for died and 0 for survived) against various covariates; the regression line can be interpreted as the probability of death by 28 days 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 $\text{MAP} < 75\text{mmHg}$), 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.

4.3.2 Determinants of 28-day mortality: an expanded role for TB therapy?

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 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[13], and all sepsis guidelines stress the importance of rapid administration of antimicrobials[14]. 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[15]. Estimates from this study are at least consistent with those, though a lack of precision here could be due to underpowering.

It is also possible, of course, that the beneficial effect of rapid antimicrobials is reduced in a population with a diverse range of causes of sepsis, or a population with a very delayed presentation to hospital, as here. There was some suggestion of a late, possibly nonlinear, deleterious effect of delay of antibacterials after around 40 hours, though confidence intervals were wide and this could represent random variability. Alternatively, antimicrobials started this late after admission could represent hospital-acquired infection, which may well confer a high mortality risk.

4.4 Discussion

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[16]. The STAMP trial[17] 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[18], 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[19] 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)[20]. 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.

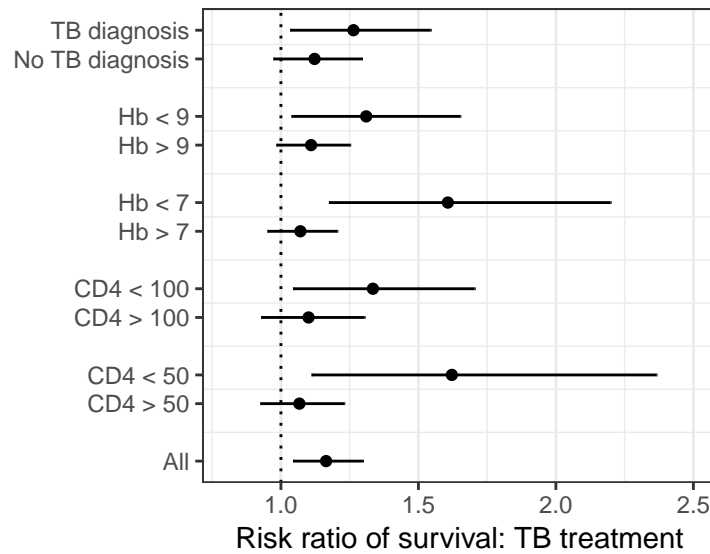


Figure 4.5: Subgroup analysis of effect of TB therapy on mortality. Crude (unadjusted) risk ratio for survival to 28 days is given; $RR > 1$ favours TB therapy, $RR < 1$ favours no TB therapy. A significant effect is seen in the immunosuppressed, anaemic, and to a lesser extent, those with a confirmed diagnosis of TB.

The WHO provides guidance on empiric TB therapy in inpatients, however[21]. Hospitalised HIV-infected patients in high TB burden settings with cough and so-called “danger signs” (fever $> 39^{\circ}\text{C}$, 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[22]. 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.

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 ($\times 10^9$ g dL⁻¹)	9.7 (7.4-11.3)	11.1 (8.6-13.9)	0.001
Severity Variables			
Temperature ($^{\circ}\text{C}$)	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 ($\times 10^9$ L ⁻¹)	6.4 (4.6-9.1)	6.6 (4.3-11.7)	0.595
Platelet count ($\times 10^9$ 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)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	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-Wallis 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

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 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
- 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 Gelman A, Carlin JB, Stern HS *et al*. *Bayesian data analysis*. 3rd ed. Chapman; Hall/CRC 2004.
- 11 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

- 12 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
- 13 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
- 14 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
- 15 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
- 16 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
- 17 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
- 18 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
- 19 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.
- 20 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.
- 21 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.
- 22 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