Chapter 4

Sepsis in Blantyre, Malawi

4.1 Chapter overview

4.2 Aims and Methods

The clincial and laboratory methods of the clinical study are described in Chapter 2, Methods; a further overview of chapter aims and description of the statistcial analysis used is given here. The aims of the analysis presented here are twofold. First, a description of the presentation, aetiology, therapies delivered, and outcomes of sepsis in Blantyre;that is, the demographics and health seeking behaviour of patients presenting with sepsis, along with description of microbiologic causes identified, therapies delivered (antimicrobials and intravenous fluids), outcomes to 180 days in terms of mortality presented at 28, 90 and 180 days, and health related quality of life. Second, an analysis to identify associations of mortality and to develop models to attempt to understand the causal effect of interventions delivered to patients presenting with sepsis.

For the first aim - description of sepsis in Blantyre - patient demographics, healh seeking behaviour, symptoms and admission physiology are described as medians and interquartile ranges for continuous variables or proportions for categrorical variables, and Kruskal-Wallace or Fisher's exact tests used to compare between groups. Aetiology is presented as simple proportions, stratified by HIV status and co-infections visualised a Venn diagrams and UpSet plots using *eulerr* and *UpSetR* packages, respectively, in R. Mortality is presented as simple proportions at 28, 90 and 180 days, and Kaplan-Meier estimation of the survival function generated using the *survival* package in R; both of these estimates are presented aggregated and stratified by HIV status, with log-rank test used to test the hypothesis that HIV-infected

and uninfected survival functions differ. Morbidity was assessed as health related quality of life (HRQoL) using the EQ-5D-3L questionaire, which assesses HRQoL across five domains (anixety, pain, self care, usual activities, walking) with participants describing their problems across the domains on a 3 point ordinal scale: no problems, moderate or extreme. So-called tariff sets are used to convert these scores to an overal utility score, which compares the health state compared to perfect heath: a utility score of one indicates that a health state is the same as usual health and zero represents death, but scores of below zero (states worse than death) are possible. Tariff sets are country specific, and no Malawian tariff sets are available, so a Zimbabwean tariff set was used. The *eq5d* package in R was used to convert health states into utility scores. HRQoL was measured at baseline, and the 1,4,12 ad 24 week visits, and is presented as proportion of participants reporting at least moderate imparment in each domain (with exact binomial confidence intervals) at each time point, as well as boxplots of utility score. Utility scores of participats with sepsis was compared to community controls using t-tests.

The second aim - models to understand the causal effect of interventions delivered to patients with sepsis - presents conceptual and technical difficulties. There are a number of standard modelling approaches in the biomedical literature when putative associations between predictor variables and mortality are to be identifed. 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 selecion 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 [Harrell book]. But selecting variables to be included in a regression model is a difficuly 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.

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 severeity (e.g physiological variables quantifying shick, 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 likley 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 wholy 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. This has been called the "Table 2 fallacy." 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. The causal inference framework provides tools to do this using DAGs.

In this chapter, therefore, the aim of the analysis is to provide an estimate of the effect of treatments administered on mortality; the class of antimicrobial administered 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. This will require correcting for (or conditioning on) host, infection, and severity variables (assuming a direct effect of infection and host on outcome). 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 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; the untransformed variables are correlated (e.g. shocked patients are likely to have elevated lactate, low blood pressure, low bicorbonate, and high heart rate) and so parameter estimates become very large when entered into a regression model. This is the problem of colinearity.

Non-normally distributed variables (temperature, white cell count, urea, creatinine, lactate) were identified on kernal 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 folowing logarithmic transformation and were recoded as binary variables with a cut off ofless 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 witha 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 without losing information). Volume of intravenous fluid administered over 6 hours was included as a linear continous predcitor in this model.

Because of nonidenifiability of the model under a maximum likelihood framework, Bayesian regression using the *brms* package in R (a front end to the Stan probabalistic programming language) 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. Four Markov-chain Monte-Carlo (MCMC) chains each with 1000 iterations and a burn-in of 500 iterations were used with default *brms* settings. Convergene 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, with each missing variable 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 all posterior parameter estimates.

Finally, the relationshp between time-to-antimicrbials and mortality was assessed, initially in bivariate associations using nonparametric locally estimates scatterpoly smoothing (LOESS) regression which performs a rolling linear regerssion 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 thre 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.

Regression difficulties and solutions

colinearity/variable selection -> FAMD nonidentifiability (zero cells) -> bayesian regression with weakly informative priors

4.3 Results

4.3.1 Study population

Figure 4.2 shows flow through the study. 225 participants were recruited in 20 months between 19th February 2017 and 2nd October 2018. Participants were recuited, in general, soon after arrival in hospital, a median (IQR) of 1.5 (0.8-2.6) hours after fist attendance. In total, 4 participants (2%) were lost to follow up over the 180-day study period; 5 participants (2%) withdrew; and 7 participants (3%) transferred out of the study area before 180 days. Four of the five participants who withdrew gave a reason for their wish to withdraw, all that they no longer wished the inconvenience of being involved in the study. 15/225 (7%) participants had their final study visit before 180 days, and so were not included in the 180-day outcome analysis.

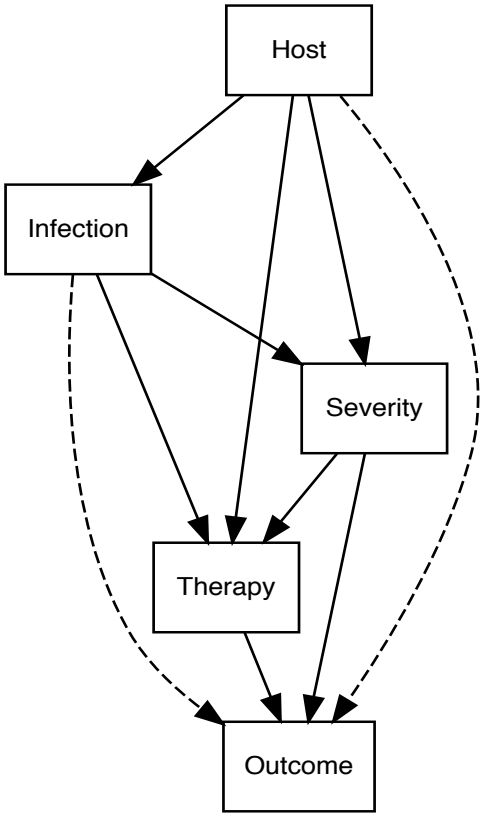

Figure 4.1: Hypothesised causal structre of death in sepsis. Host factors (e.g. age, sex, immune status) influence the type of infection; disseminated TB is more common in HIV, for example. Severity (variables quantifying e.g. shock or respiratory failure) is influenced by infection type and host factors. 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

4.3.2 Symptoms and health-seeking behaviour

Table 4.1 shows the baseline characteristics of the recruited participants. They were young (median [IQR] age 36 [28-44]) and predominantly HIV-infected. Of those who were HIV-infected, the majority (117/143 [82%]) were on ART, almost exclusively the Malawian first-line regimen of efavirenz, lamivudine and tenofovir, and 88/117 (75%) had been taking ART for more than three months. Figure 4.3 shows the presenting symptoms of the participants. Almost all (221/225 [98%] of participants) experienced subjective fever. Participants had been unwell for some time, a median (IQR) of 7 (3-14) days; 32/225 (14%) of participants had been unwell for more than 4 weeks. 18/225 (8%) of participants had been admitted to hospital within the last 4 weeks. Over half (123/225 [55%]) of participants had sought care for their current ilness (Table 4.2), most commonly (101/123 [82%] of participants) at a government health centre, a median (IQR) of 2 (1-6) days previously. 60/225 (27%) of all participants had recieved an antimicrobial for their current illness: 7/60 (12%) of all prehospital antimicrobials were antimalarials, the remainder anitbacterial, most commonly co-trimoxazole or ciprofloxacin. Prehopsital intravenous or intramuscular antimirobials were administered in 16/60 (27%) participants recieving antimicrobials: ceftriaxone (n=6), benzylpenicillin (n=4), gentamicin (n=3) and artesunate (n=3).

Table 4.1: Participant Characteristics

Variable	Value
Demographics	
Age (years)	36 (28-44)
Male sex	114/225 (51%)
HIV/TB status	
HIV Reactive	143/225 (64%)
HIV Non Reactive	70/225 (31%)
HIV Unknown	12/225 (5%)
Ever treated for TB	37/225 (16%)
Of those, current TB treatment	10/37 (27%)
ART status*	
Current ART	117/143 (82%)
Months on ART	29 (4-73)
ART regimen: EFV/3TC/TDF	110/117 (94%)
ART regimen: other	7/117 (6%)
Current CPT†	98/141 (70%)
Tobacco/alcohol use	
Never tobacco	196/225 (87%)
Ex tobacco	17/225 (8%)
Current tobacco	12/225 (5%)

Table 4.1: Participant Characteristics *(continued)*

Variable	Value
Current alcohol	51/225 (23%)
Education	
Primary incomplete or complete	97/225 (43%)
Secondary school complete	48/225 (21%)
Some secondary education	47/225 (21%)
College or higher	17/225 (8%)
No formal schooling	16/225 (7%)
Employment	
Unemployed	82/225 (36%)
Currently employed	65/225 (29%)
Self-employed	56/225 (25%)
Student	21/225 (9%)
Retired	1/225 (0%)
Toilet facilities	
Pit latrine with slab +/- foot rest	104/225 (46%)
Hanging toilet/latrine	59/225 (26%)
Pit latrine with slab and cover +/- foot rest	45/225 (20%)
Flush Toliet (any type)	14/225 (6%)
No toilet	2/225 (1%)
Composting toilet	1/225 (0%)
Main water source	
Piped outside dwelling	69/225 (31%)
Tube well/borehole	64/225 (28%)
Public tap/standpipe	51/225 (23%)
Piped into dwelling	30/225 (13%)
Unprotected well/spring	5/225 (2%)
Surface water (including rainwater collection)	4/225 (2%)
Tube well with powered pump	2/225 (1%)
Electricty	
Electricity available in house	119/225 (53%)
Main cooking fuel	
Charcoal	161/225 (72%)
Wood	61/225 (27%)
Electricity	3/225 (1%)
Animals at home?	
Any animal	71/225 (32%)
Poultry	46/71 (65%)
Dogs	18/71 (25%)
Goats	12/71 (17%)
Dogs	18/71 (25%)
Other	11/71 (15%)

Table 4.1: Participant Characteristics *(continued)*

Variable	Value

Note:
ART = Antiretroviral therapy, CPT = Co-trimoxazole preventative therapy, EFV: Efavirenz, 3TC: Lamivudine, TDF: Tenofovir. Numeric values are median (IQR)) unless otherwise stated.

* ART status includes HIV reactive only as denominator

† Missing CPT data for two participants.

Table 4.2: Prehospital heathcare seeking and antimicrobial exposure

Variable	Value
Pre-hospital healthcare seeking	
Sought care prior to attendance at hospital	123/225 (55%)
At health centre	101/123 (82%)
At hospital	16/123 (13%)
At private doctor	8/123 (7%)
Somewhere else	1/123 (1%)
Days prior to today that participant sought care	2 (1-6)
Prehospital antimicrobial exposure	
Recieved any antimicrobial prior to attendance at hospital	60/225 (27%)
Co-trimoxazole	12/60 (20%)
Ciprofloxacin	10/60 (17%)
Amoxicillin	9/60 (15%)
Ceftriaxone	6/60 (10%)
Metronidazole	5/60 (8%)
Benzylpenicillin	4/60 (7%)
Artesunate	3/60 (5%)
Gentamicin	3/60 (5%)
Erythromycin	2/60 (3%)
LA	2/60 (3%)
SP	2/60 (3%)
Azithromycin	1/60 (2%)
Flucloxacillin	1/60 (2%)
Days prior to today that antimicrobials started	2 (1-5)
Method of transport to hospital	
Minibus	78/225 (35%)
Taxi	65/225 (29%)
Private car/truck	42/225 (19%)
Ambulance	37/225 (16%)
Other	2/225 (1%)
Walk	1/225 (0%)
Cost (MWK) of transport to hospital	1000 (275-3000)

Note:
LA = Lumefantrine-artemether, SP = Sulfamethoxazole-pyrimethamine, MWK = Malawian Kwacha. Numeric values are median (IQR)) unless otherwise stated.

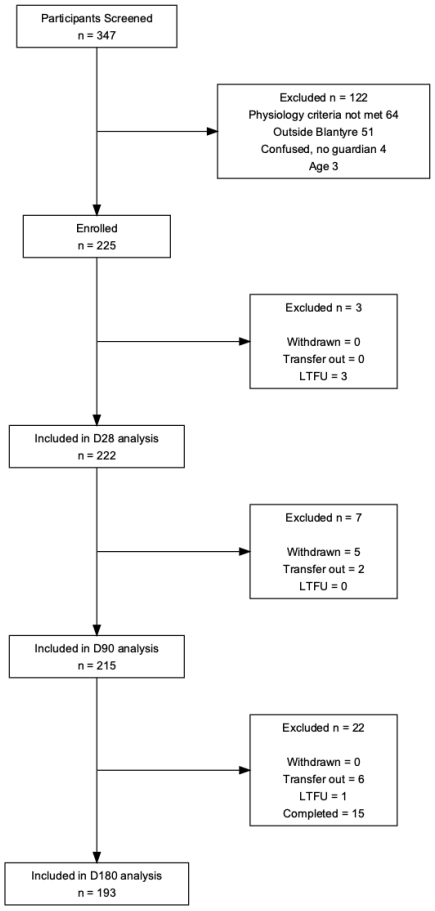

Figure 4.2: Study recruitment and follow up.

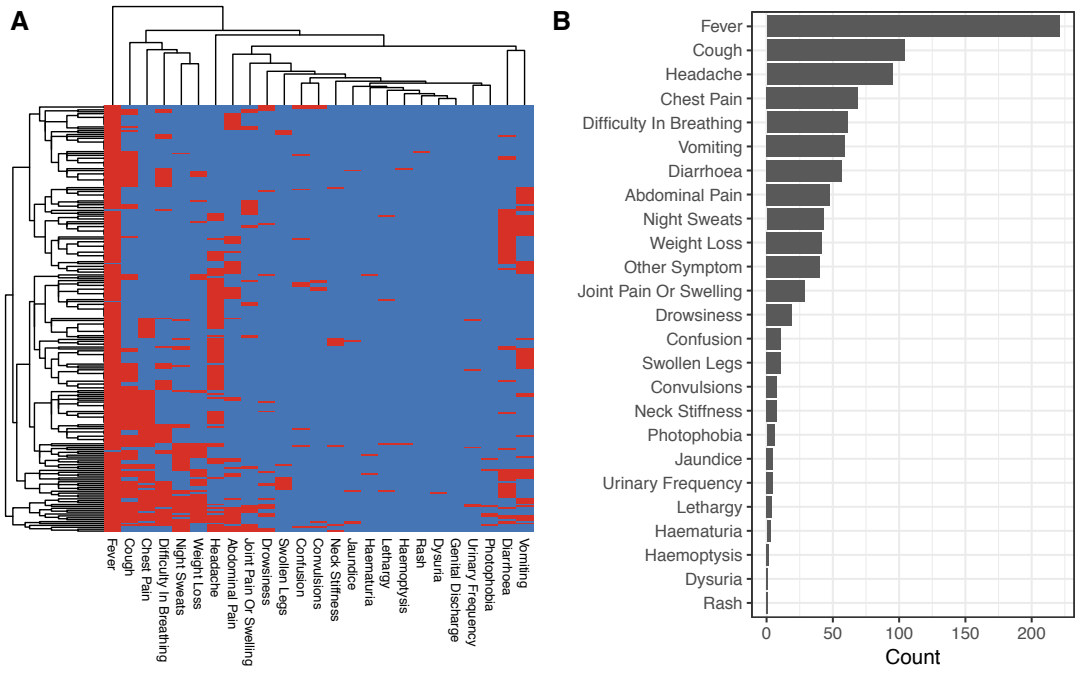

Figure 4.3: Symptoms of recruited participants. A: Row and column clustered heatmap of participant symptoms. Each row represents a patient. Red = presence, blue = absence. B: Frequency of occurence of symptoms

4.3.3 Admission physiology and laboratory investigations

Admission vital signs and laboratory investigations are shown in Table 4.3. Despite high ART coverage (117/143 [82%]) amongst HIV-infected participantsfor a median of 29 months, the median (IQR) CD4 count was low at 98 (31-236) cells μL^{-1}. 108/141 (70%) of participants had a CD4 count below 200 cells μL^{-1}. CD4 count was similar in participants who had started ART more than 6 months ago as compared to less than three months ago (median [IQR] 99 [27-260] vs 93 [39-137] cells μL^{-1} respectively) and 42/83 (51%) of participants who had been taking ART for more than 6 months had a CD4 count of less than 100 cells μL^{-1}, and would fulfil a WHO definition of immunological failure.

Table 4.3: Admission physiology, haematology and biochemistry

Variable	Value
Admission physiology	
Temperature (°C)	38.5 (37.9-39.0)
Heart rate (min^{-1}))	121 (102-132)
Systolic BP (mmHg)	99 (85-119)
Diatsolic BP (mmHg)	66 (56-76)
MAP (mmHg)	76 (65-89)
Respiratory rate (min^{-1})	34 (32-38)
Oxygen saturation (%)	96 (94-98)
GCS 15	204/225 (91%)
GCS 11-14	16/225 (7%)
GCS < 11	5/225 (2%)
Unable to stand	63/225 (28%)
Admission CD4 count	
CD4 count* (μL^{-1})	98 (31-236)
Admission haematology	
Haemoglobin ($x10^9$ g dL^{-1})	10.8 (8.2-13.2)
White cell count ($x10^9$ L^{-1})	6.5 (4.4-11.4)
Neutrophil count ($x10^9$ L^{-1})	4.0 (2.1-7.5)
Platelet count count ($x10^9$ L^{-1})	218 (146-297)
Admission biochemistry	
Sodium (mmol L^{-1})	134 (130-137)
Potassium (mmol L^{-1})	4.0 (3.6-4.4)
Bicarbonate (mmol L^{-1})	19 (17-22)
Chloride (mmol L^{-1})	101 (97-104)
Urea (mmol L^{-1})	4.8 (3.5-8.0)
Creatinine (mmol L^{-1})	76 (59-103)
Lactate (mmol L^{-1})	3.4 (2.3-5.2)

Note:
GCS = Glasgow coma scale, BP = Blood pressure, MAP = Mean arterial blood pressure. Numeric values are median (IQR)) unless otherwise stated.

* CD4 count includes only HIV-infected participants; 2 values were missing.

4.3.4 Aetiology

In total, 51% (114/225) of the 225 participants had at least one infectious agent identified (Table 4.4), most commonly tuberculosis (76/225 [34%]) followed by bloodstream infection (24/225 [11%]) and malaria (21/225 [9%]). Table 4.5 shows the availability of test and proportion of positive tests across the cohort, stratified by HIV status. 2/225 patients (1%) had a missing aerobic blood culture; the remaining 223 patients had a total of 259 blood cultures performed. 15/259 (6%) blood cultures grew at least one contaminant, but 26 blood cultures from 24 patients were positive for a total of 28 pathogenic bacteria (Figure 4.4): *Salmonella* Typhi was the most commonly isolated pathogenic bacterium, and seemed to show an association with HIV-negative participants: all (8/8) of the participants from whom *S.* Typhi was isolated and whose HIV status was known were HIV noninfected. Of the 18 Gram negative bacteria isolated, 3/18 (17%) were cefpodoxime resistant on AST via disc diffusion testing, and likely ESBL producers: one *K. pneumoniae* and one *E. coli* (both from the same blood culture and same patient) and one *Acinetobacter baumannii*. Both *Staphyloccus aureus* isolates were oxacillin sensitive. The one *Streptococcus pneumoniae* cultured was penicillin intermediate on AST.

Lumbar puncture and CSF culture was carried out in 44 participants: 5/44 (11%) of samples grew a contaminent and no pathogenic bacteria were recovered from any sample. 4/44 (9%) had a detectable cryptococcal antigen (CRAG) in CSF. Malaria testing was missing for 6/225 (3%) of participants, but of the remainder, a positive malaria test was more likely in the HIV-uninfected (12/69 [17%] vs 6/138 [4%], p = 0.01 on pairwise Fisher's exact test). Positive aerobic blood culture showed no statistically significant association with HIV, nor did positive CSF testing, though in the latter case numbers were small and all positive tests (all positive CRAG) were in fact in the HIV-infected (Table 4.5).

Testing for TB, with the exception of sputum Xpert testing, was restricted to HIV-infected participants. Sputum Xpert was carried out in 44/225 (20%) of participants, and was more commonly carried out in the HIV-infected: 35/143 [24%] of HIV-infected participants had sputum testing performed vs 8/70 (11%) of HIV uninifected (p = 0.07 by Fisher's exact test). 53 sputum samples were sent in total from the 44 patients, and 8/44 (18%) diagnoses of TB made, all except one in HIV-infected participants. One sample identified a rifampicin resistance gene; the remainder of infections were rifampcin-sensitive.

155 participants were eligible for urinary lipoarabinomannan (uLAM) and mycobacterial blood culture testing, being either HIV-infected (n=143) or of unknown HIV status (n=12). Urine was available for 145/155 (94%) of those eligible, and 74/145 (51%) of samples were positive for uLAM. 150/155 (97%) of eligible participants had blood samples collected and cultured

Table 4.4: Final diagnosis of all participants

Diagnosis	Proportion of participants
Tuberculosis	76/225 (34%)
Bloodstream infection	24/225 (11%)
Malaria	21/225 (9%)
Meningitis	4/225 (2%)
No diagnosis	111/225 (49%)

Table 4.5: Positive diagnostic tests for all participants, stratified by HIV status.

	HIV status				
Test	Positive	Negative	Unknown	**All**	p
Number of participants	143	70	12	**225**	-
TB diagnostics					
Urinary LAM	70/136 (51%)	-	4/9 (44%)	**74/145 (51%)**	-
Sputum Xpert	7/35 (20%)	1/8 (12%)	0/1 (0%)	**8/44 (18%)**	0.835
TB blood culture	7/128 (5%)	-	1/10 (10%)	**8/138 (6%)**	-
Other diagnostics					
Aerobic blood culture	13/141 (9%)	9/70 (13%)	2/12 (17%)	**24/223 (11%)**	0.647
CSF culture or CRAG	4/31 (13%)	0/12 (0%)	0/1 (0%)	**4/44 (9%)**	0.445
Malaria RDT	6/138 (4%)	12/69 (17%)	3/12 (25%)	**21/219 (10%)**	0.007

Note:
LAM = Lipoarabinomannan, CSF = Cerebrospinal fluid, CRAG = Cryptococcal antigen, RDT = Rapid diagnostic test. p-values are chi-squared test across the three HIV status strata, and hence may be different from the pairwise exact Fisher's tests presented in the text. Urinary LAM and TB blood culture were not carried out in HIV negative participants.

for mycobacteria. 12/150 (8%) grew contaminents and are excluded from the denominators in Table 4.5; of the remainder 8/138 (6%) grew mycobacteria, all *M. tuberculosis*.

Figures 4.6 and 4.5 show the overlap of positive tests form the different diagnostic modalities. Of the 114 patients with at least one positive diagnosic test, 90/114 (79%) had only one positive diagnostic test. The exceptions to this were mycobacterial blood culture and sputum Xpert: patients who had TB diagnosed by these tests tended to also have a positive uLAM. 2/4 (50%) of patients with positive CSF testing (all of whom had detectible CRAG) had also grew *Cryptococcus neoformans* in aerobic blood culture. 111/225 (49%) of patients remained with no diagnosis.

4.3.5 Treatment

At least one antimicrobial drug was recieved by 95% (214/225) of the cohort during their admission (Table @ref:(time-to-ab-table)), most commonly an antibacterial (207/225 [92%]), but also a significant minority recieved antitubercular therapy (63/225 [28%]). Of those

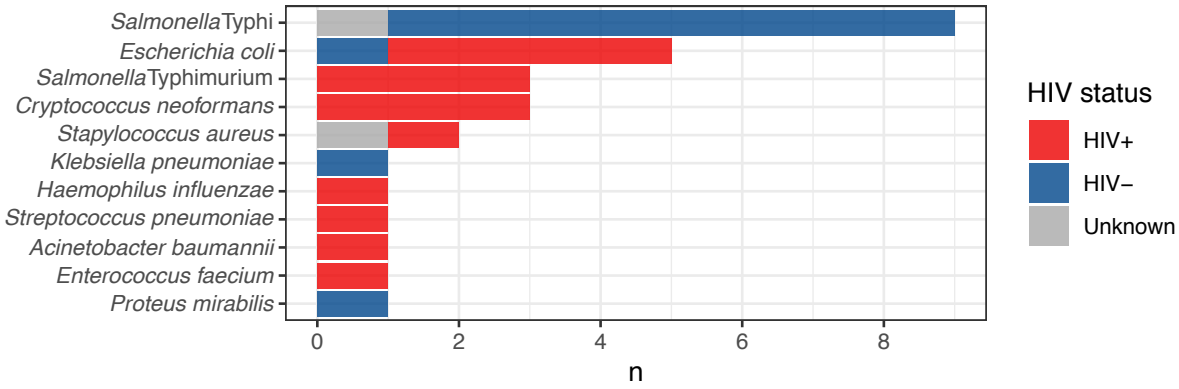

Figure 4.4: Pathogenic isolates recovered from aerobic blood culture. 26 blood cultures in 24 participants were positive for 28 pathogens in total

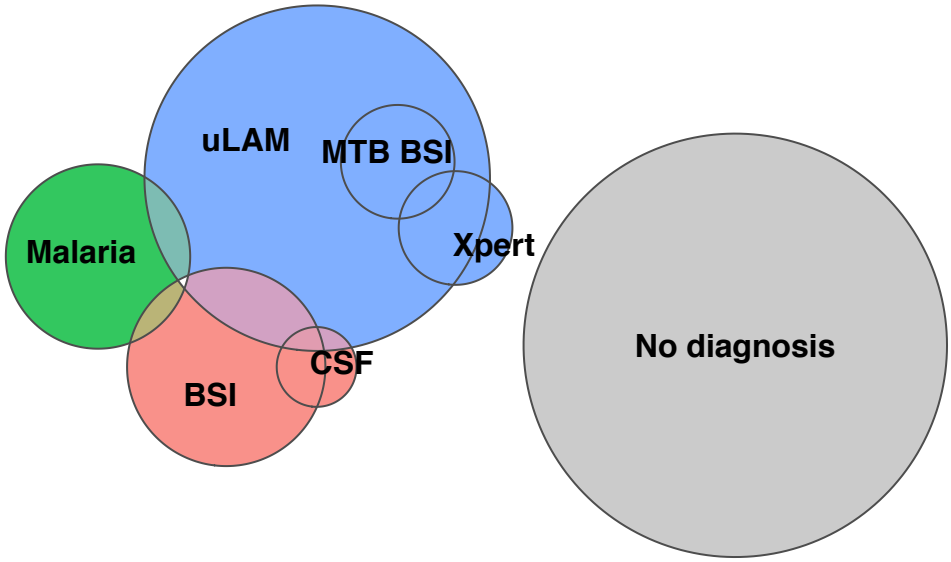

Figure 4.5: Venn diagram showing overlap of positive diagnostic tests; culture of blood and CSF shown in red, malaria in green and TB diagnostics in blue. The CSF variable in includes either a positive culture for a pathogenic bacteria or positive cryptococcal antigen, BSI a positive aerobic culture of pathogenic bacteria from blood and MTB BSI a positive mycobacterial culture of tuberculosis from blood. BSI: Bloodstream infection, CSF: Cerebrospinal fluid, CRAG: Cryptococcal antigen, mRDT: Malaria rapid diagnostic test, MTB BSI: Mycobacterium tuberculosis bloodstream infection, uLAM: urinary lipoarabinomannan.

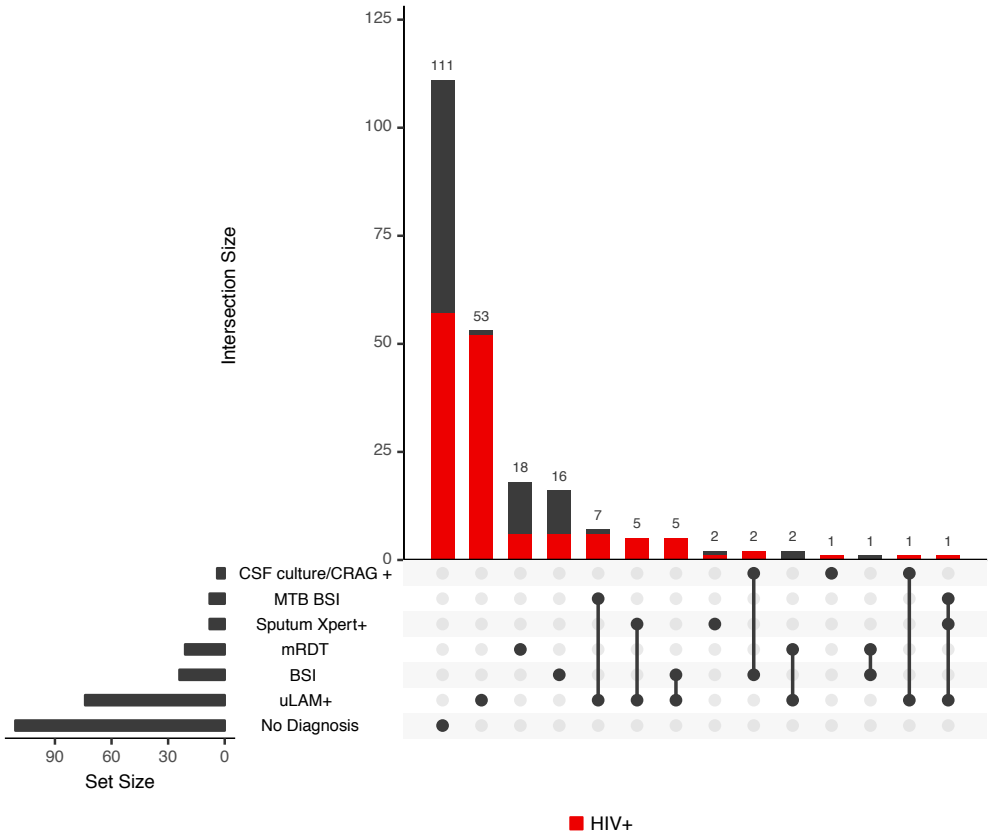

Figure 4.6: UpSet plot of overlap of positive diagnostic tests, showing that for the majority of participants, one test alone is positive. Red colour indicates HIV-infected; black is a composite of HIV-negative and unknown. The CSF variable in includes either a positive culture for a pathogenic bacteria or positive cryptococcal antigen, BSI a positive aerobic culture of pathogenic bacteria from blood and MTB BSI a positive mycobacterial culture of tuberculosis from blood. BSI: Bloodstream infection, CSF: Cerebrospinal fluid, CRAG: Cryptococcal antigen, mRDT: Malaria rapid diagnostic test, MTB BSI: Mycobacterium tuberculosis bloodstream infection, uLAM: urinary lipoarabinomannan.

Table 4.6: Door-to-antimicrobial times.

Antimicrobial class	No. participants	Median [IQR] time (hours)
Antibacterial	207/225 (92%)	5.3 (3.7-10.8)
Antitubercular	63/225 (28%)*	120.9 (63.7-171.0)
Antifungal	26/225 (12%)	47.7 (27.9-73.9)
Antimalarial	12/225 (5%)	4.5 (3.1-21.7)

* 10/63 participants who received antitubercular agents during admission were taking them prior to admission; they are excluded from the calculation of median door-to-antimicrobial time for this class.

receiving antitubercular therapy, 16% (10/63) were taking the medication prior to admission, and the remainder were initiated on therapy during admission. The first antibacterial agent administered was most often ceftriaxone, in 87% (181/207) of cases but ciprofloxacin (18/207 [9%] of participants), amoxicillin (6/207 [3%]) and metronidazole (2/207 [1%]) were also used. Median door to antimicrobial time was 5.3 (IQR 3.7-10.8) hours for antibacterials and 4.5 (IQR 3.1-21.7) hours for antimalarials but longer for antifungals at 47.7 (IQR 27.9-73.9) hours and longer still for antitubercular therapy at 120.9 (IQR 63.7-171.0). Cumulative incidence curves for administration of the different antimicrobial classess are shown in Figure 4.7A-D.

Of all participants, 85% (192/225) received any intravenous fluid in the first 6 hours of enrollment to the study; of these, most received 0.9% saline (160/192 [83%] of those recieving fluid) but 5% dextrose (91/192 [57%]) were also used; Ringer's lactate (6/192 [6%]) and blood (2/192 [1%]) were rarely administered. Of the 192 patients who were administered any fluid, a median of 1.5L (IQR 1-2L) was administered over the 6hr study period; fluid administration as a function of time is shown in Figure 4.7E.

4.3.6 Outcome

Median hospital stay was 4 (IQR 1-9) days. Mortality of the cohort was 18% (95% CI 13-23%) at 28 days, 24% (95% CI 18-30%) at 90 days and 31% (95% CI 25-38) at 180 days, and higher in HIV-infected participants at each time point (Table 4.7), though not statistically significant on pairwise Fisher's exact test (HIV-infected vs noninfected 19% vs 13%, [p = 0.14] at 28 days, 27% vs 17%, [p = 0.44] at 90 days and 36% vs 21% [p = 0.29] at 180 days). Kaplan-meier estimation of the survival function (Figure 4.8) showed a precipitous decline in survivorship to around day 30 and mortality at a reduced rate thereafter, to the end of the study period. Stratifying the analysis by HIV status revealed that early deaths (within the first 1-2 weeks) occur at similar rates in the two groups before the curves diverge; log-rank test suggested a significant difference in survival function between the two groups (p = 0.03).

Health related quality of life measures, as assessed by EQ-5D-3L, are shown in Figure 4.9 for

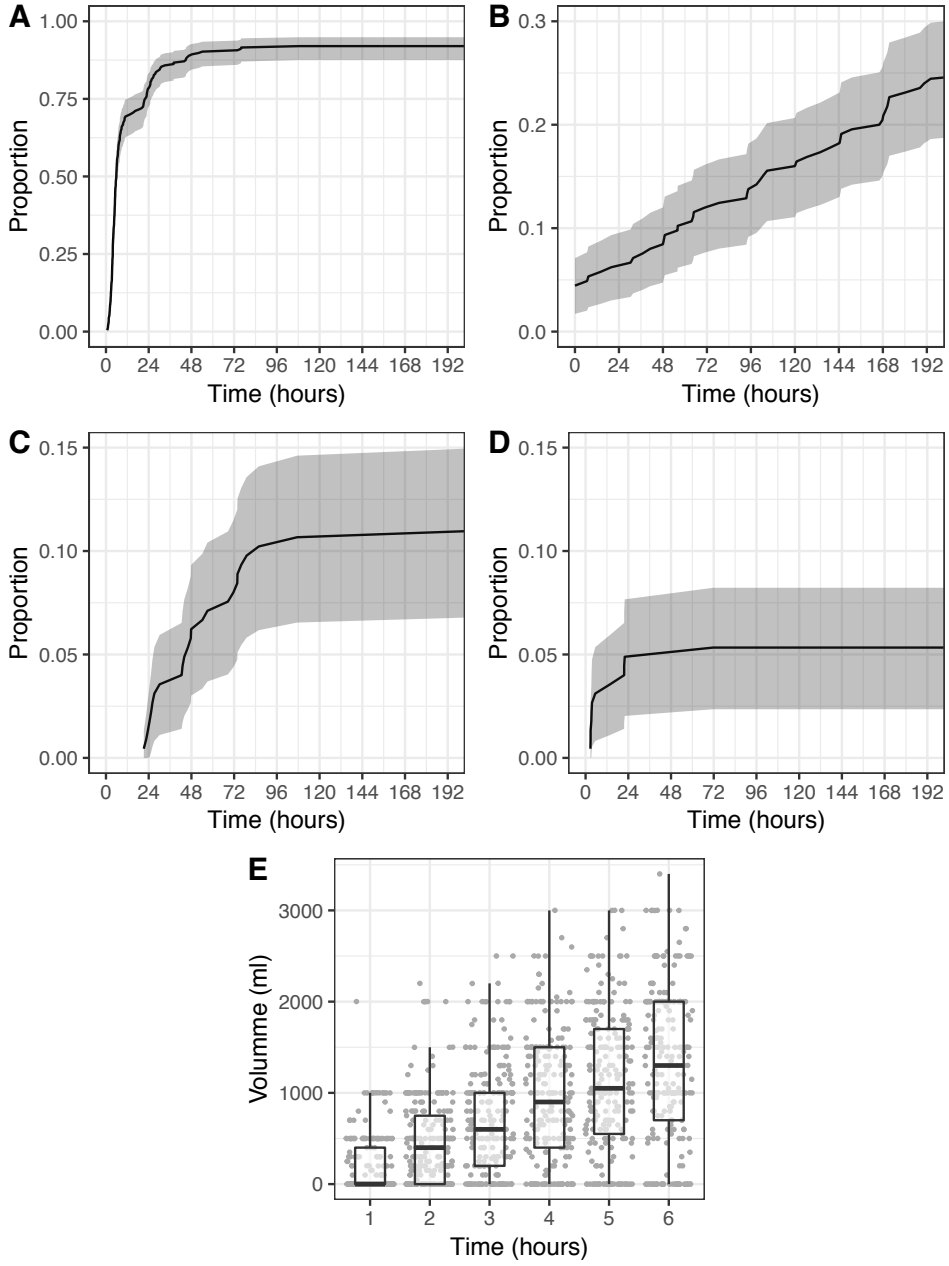

Figure 4.7: Timing of antimicrobial and fluid administration. A-D: Cumulative incidence of administration of antibacterial (A), antitubercular (B), antifungal (C) and antimalarial (D) agents as a function of time since arrival at hospital in hours. E: Total volumne of administered intavenous fluid as a function of time since enrollment to study in hours. Boxplots show median, quartiles box and 1.5 times interquartile range as whiskers. Points are jittered around the hour at which they were measured to show distribution.

Table 4.7: Day 28, 90 and 180 mortality stratified by HIV status

	HIV+		HIV-		HIV Unknown		Total	
	n	Mortality	n	Mortality	n	Mortality	**n**	**Mortality**
Day 28	143	19% (13-26)	67	13% (6-24)	12	25% (5-57)	**222**	**18% (13-23)**
Day 90	139	27% (19-35)	64	17% (9-29)	12	25% (5-57)	**215**	**24% (18-30)**
Day 180	125	36% (28-45)	58	21% (11-33)	11	27% (6-61)	**194**	**31% (25-38)**

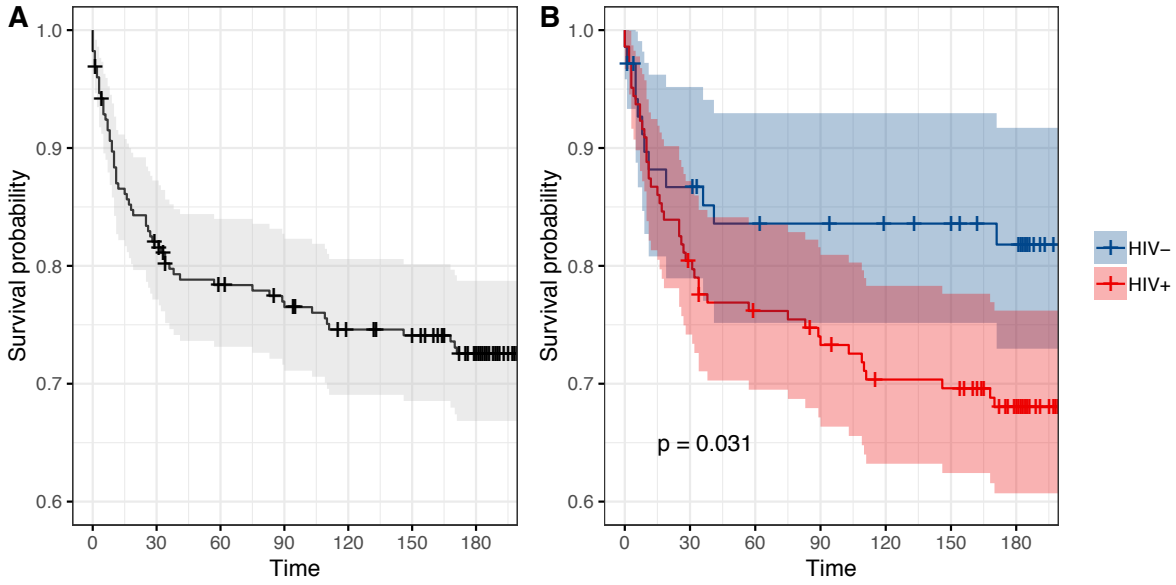

Figure 4.8: Kaplan-Meier survival curves of all included participants (A) and stratified by HIV status (B). Crosses indicate censoring. p value from log-rank test comparing survival of HIV-infected to HIV-noninfected participants shown (p = 0.03).

participants with sepsis and the community cohort as a comparator. Acutely, participants with sepsis reported were significantly disabled, reporting at least moderate impairment across all domains in the majority of cases, and over 90% of participants reporting at least moderate impairment in activities of daily living and experienceing at least moderate pain or discomfort. However, recovery following treatment in survivors was rapid. The mean EQ-5Q utility score (a measure of the weight compared to a health state compared to 1, perfect health) of healthy community controls was 0.910 (SD 0.102) at enrollment, significantly higher than participants with sepsis at enrollment (utility score 0.496 (SD 0.251), p = $<$ 0.0001 versus community scores by t-test), but comparable to participants with sepsis at their 12 week assessment (0.913 (SD 0.147), p = 0.903 versus community enrollment scores).

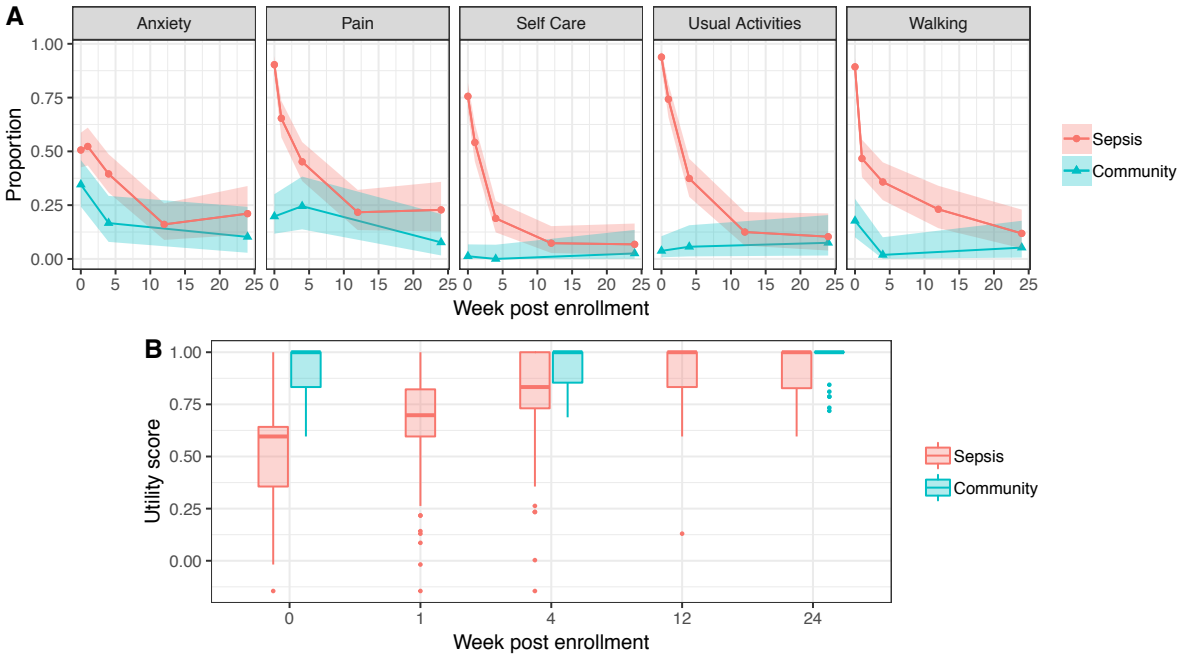

Figure 4.9: Health-related quality of life following sepsis admission, compared to community controls, showing a return to usual quality of life by 12 weeks following admission. A: proportion of participants across each of the five domains of the EQ-5D questionnaire who report at least moderate impariment. B: EQ-5D utility score derived using the Zimbabwean tariff set. The utility score is interpreted as the weight attached to a health state compared to perfect health, which is assigned a value of 1. By 12 weeks there is no statistically significant difference between sepsis and baseline community participant utility scores (p = 0.90 by t-test).

4.3.7 Determinents of mortality

Bivariate associations of mortality are shown in 4.8 with variables grouped into putative host, severity, diagnosis and treatment variables. Variables associated with immunosupression - CD4 count and haemoglobin - were associated with death in bivariate analysis, as were well recognised markers of disease severity: shock, hypoxia, reduced conscious level, hyperlacateaemia, and inability to ambulate, as were reduced venous bicarbonate and increased venous urea. A diagnosis of malaria was strongly associated with survival; none of the 21 participants with a diagnosis of malaria died. Conversely, a diagnosis of meningitis was associated with mortality (Table 4.8 and Figure 4.10) though numbers were small (n = 4) and so estimates of effect size have wide confidence intervals. Reciept of antibacterials, antifungals or antimalarials showed no association with mortality, though almost all participants recieved antibacterials and only a minority antimalarials and antifugals, so moderate effect sizes would be unlikely to be detetced. However, reciept of antitubercular therapy was associated with survival: 8% (4/53) of participants recieving TB therapy died compared to 21% (35/169) who did not receive it (p = 0.04 by Fisher's exact test).

Kruskal-Wallace 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 4.10, 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 recieved antimalarial therapy died. No relationship is apparent , thanks to wide confidence intervalsexcept possibly for antibacterial therapy (Figures 4.10D and E): any effect would seem to be only apparent on delay on antimicrobial administration beyond around 40 hours post-admission, with likley non-linearity in effect size as a function of antimicrobial delay. Similarly, volume of intravenous fluid administered does has no apparent effect on 28 day mortality (Figure 4.10B). 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 4.10C) once again showed no appaent 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 kernal denisty 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 categoriacal 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 **??**C) and dimensions one and two provide some discrimination in terms of diagnosis, particularly bewteen TB and malaria (Figure **??**D 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 above. Parameter estimates and 95% credible intervals from this model are shown in 4.9, 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.

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 follwoing imputation of missing data as before. The estimates of the coefficients of these models are shown in Table 4.10 and the predicted probability of death by 28 days shown in Figure (Figure 4.12. 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.

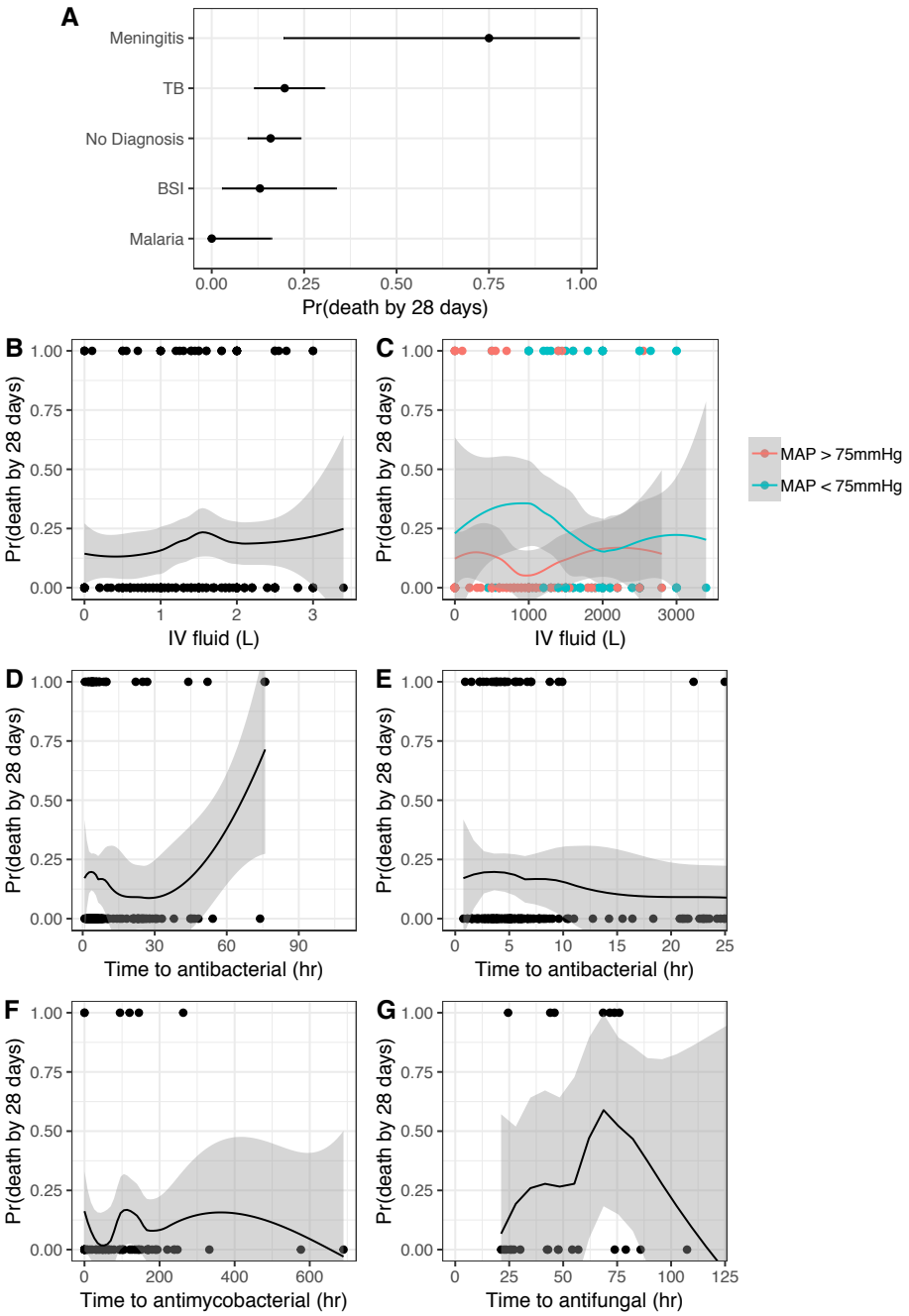

Figure 4.10: Bivariate associations of death by 28 days. A: 28-day mortality stratified by diagnosis. B-G show nonparametric regression (LOESS moving linear regression) of outcome (with death coded as 1 for died and 0 for survived) against various covariates; the regression line can be interpreted as 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. B: IV fluid (L), C: IV fluid stratified by presence or absence of shock (defined as MAP $<$ 75mmHg), D: Time to antibacterials, E: Time to antibacterials with axes restricted to 24hrs to show lack of apparent relationship during first day of admission, F: Time to antimycobacterials, G: Time to antifungals. No plot is shown for antimalarials as no participant receiveing antimalarials died. Only D seems to show a relationship, which may be nonlinear.

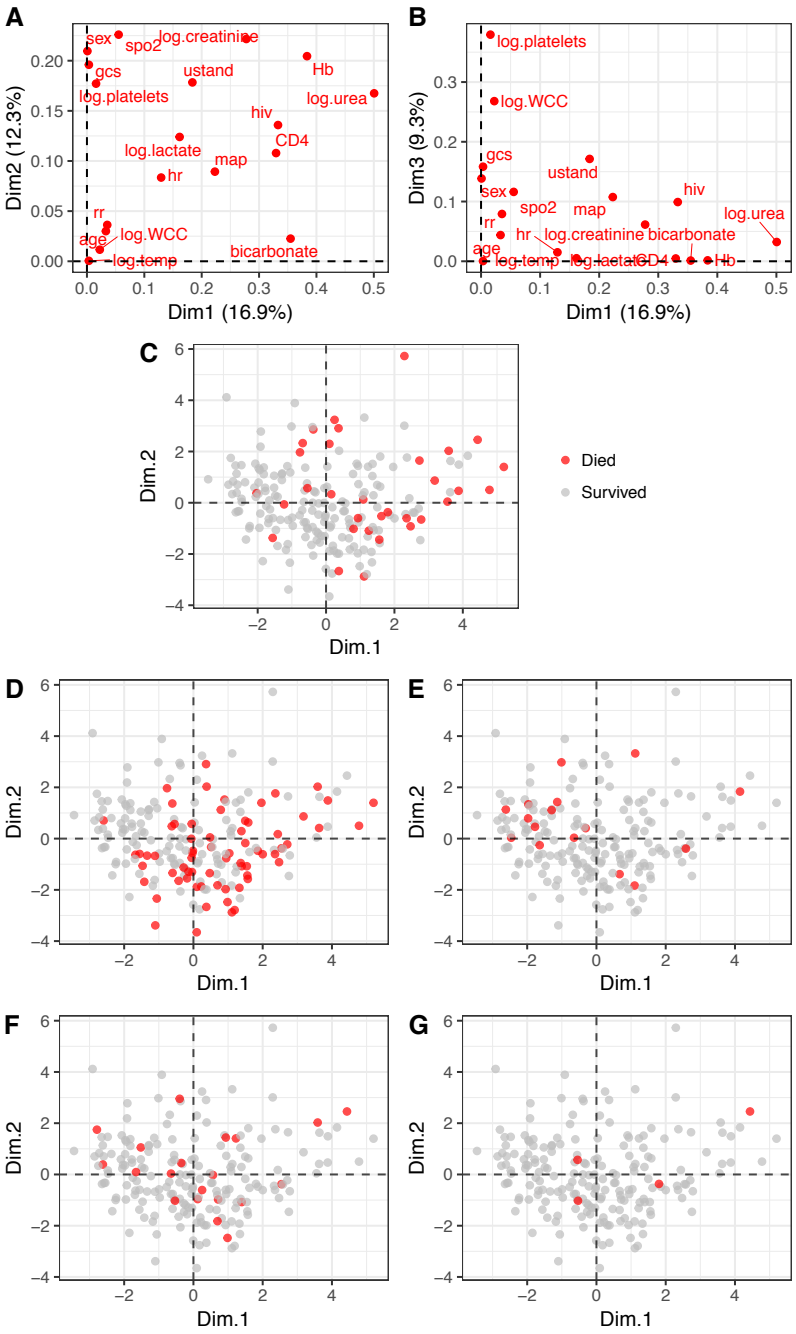

Figure 4.11: Dimensionality reduction of dataset using factor analysis of mixed data (FAMD); this is a combination of principal components analysis (PCA) for continuous variables and multiple correspondence analysis (MCA) for categorical variables, resulting in a new orthogonal coordinate system which 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 variabce 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 is associated with mortality. D-G show individuals colored by diagnosis: red in each case corresponds a diagnosis of TB (D), malaria (E), BSI (F) and meningtis (G) to show that malaria and tb seperate somewhat in Dim.1 and 2 with malaria patients clustering in top left and TB patients in bottom right

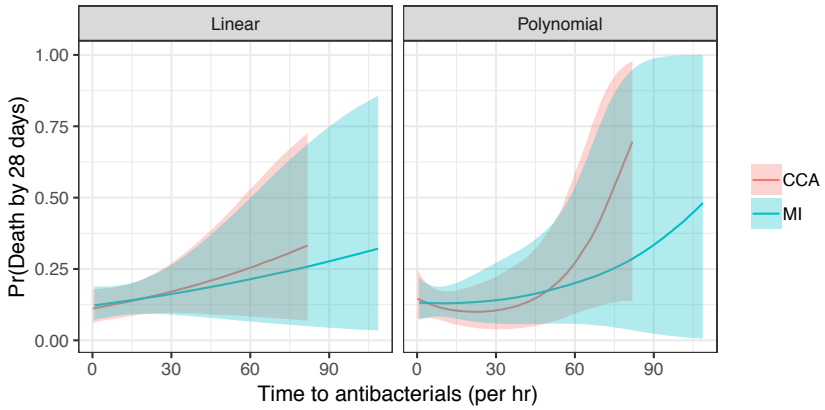

Figure 4.12: Relationship of time-to-antibacterials and 28-day mortality, adjusted for FAMD dimensions 1 and 2. Lines show predicted probability of death from modelling time to antibacteial as a linear predictor of mortality (left panel) and as a second order polynomial (right panel) in both complete case analysis (CCA) and pooled multiply imputed (MI) datasets (n=10). In each case there is no clear relationship thanks to uncertainty in the estimate.

Table 4.8: Bivariate associations with death by 28 days

Variable	Died	Survived	p
Host Variables			
Age (years)	36.4 (31.5-46.0)	35.9 (27.4-42.9)	0.252
Male sex	19/39 (49%)	93/183 (51%)	0.861
HIV Infected*	27/36 (75%)	116/174 (67%)	0.433
Taking ART†	21/27 (78%)	96/116 (83%)	0.582
CD4 count† (μL^{-1})	**28.5 (9.5-124.5)**	**103.0 (43.5-251.0)**	**0.007**
Haemoglobin (x10^9 g dL^{-1})	**9.1 (6.0-10.4)**	**11.0 (8.6-13.4)**	**<0.001**
Severity Variables			
Temperature (°C)	**38.1 (37.7-38.8)**	**38.5 (38.0-39.0)**	**0.024**
Heart rate (min^{-1}))	123.0 (104.5-138.5)	120.0 (102.0-131.0)	0.510
Systolic BP (mmHg)	**89.0 (76.0-106.0)**	**99.0 (86.5-118.5)**	**0.047**
Diastolic BP (mmHg)	**59.0 (51.0-72.0)**	**67.0 (57.0-75.5)**	**0.040**
Mean arterial BP (mmHg)	**69.7 (60.0-81.3)**	**78.7 (67.0-89.2)**	**0.035**
Respiratory rate (min^{-1}))	34.0 (32.0-36.5)	34.0 (32.0-38.0)	0.720
Oxygen saturation (%)	**95.0 (89.5-97.0)**	**97.0 (95.0-98.0)**	**0.019**
GCS	**15.0 (15.0-15.0)**	**15.0 (15.0-15.0)**	**0.044**
Unable to stand	**27/39 (69%)**	**36/183 (20%)**	**<0.001**
Lactate (mmol L^{-1})	**4.6 (3.0-10.6)**	**3.2 (2.1-4.5)**	**0.001**
White cell count (x10^9 L^{-1})	5.9 (3.5-11.0)	6.9 (4.5-11.5)	0.165
Platelet count (x10^9 L^{-1})	181.5 (86.8-300.8)	223.0 (148.0-296.5)	0.291
Bicarbonate (mmol L^{-1})	**17.0 (14.0-21.0)**	**20.0 (17.0-22.0)**	**0.007**
Urea (mmol L^{-1})	**7.8 (4.5-14.3)**	**4.5 (3.2-7.0)**	**<0.001**
Creatinine (mmol L^{-1})	90.0 (60.0-185.0)	73.0 (59.0-96.0)	0.100
Diagnosis			
BSI	3/39 (8%)	20/183 (11%)	0.773
TB	15/39 (38%)	61/183 (33%)	0.579
Malaria	**0/39 (0%)**	**21/183 (11%)**	**0.030**
Meningitis	**3/39 (8%)**	**1/183 (1%)**	**0.018**
No diagnosis	21/39 (54%)	88/183 (48%)	0.598
Treatment Recieved			
Antibacterials	37/39 (95%)	167/183 (91%)	0.746
Time to Antibacterials (hr)	4.7 (3.8-8.8)	5.3 (3.6-10.8)	0.648
Antifungals	7/39 (18%)	19/183 (10%)	0.180
Time to Antifungals (hr)	68.5 (45.0-72.7)	47.6 (26.6-76.4)	0.665
Antimalarials	0/39 (0%)	12/183 (7%)	0.132
Time to Antimalarials (hr)	NA (NA-NA)	4.5 (3.1-21.7)	NA
Antimycobacterials	**4/39 (10%)**	**49/183 (27%)**	**0.037**
Time to Antimycobacterials (hr)	107.3 (23.6-138.7)	99.0 (37.0-169.4)	0.778
IV fluid (ml)	1450.0 (1000.0-2000.0)	1300.0 (625.0-2000.0)	0.368

Note:
BP = Blood pressure, GCS = Glasgow coma scale. Numeric variables are presented as median (IQR) and categoric 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

Table 4.9: Adjusted odds ratios (aOR) for death by 28 days following multiple imputation of missing data

Variable	aOR
FAMD composite variables	
FAMD Dimension 1	2.53 (1.82-3.70)
FAMD Dimension 2	1.30 (0.97-1.76)
FAMD Dimension 3	1.09 (0.78-1.53)
Diagnosis	
TB	0.72 (0.27-1.83)
Malaria	0.03 (0.00-0.76)
BSI	0.25 (0.04-1.15)
Meningitis	16.18 (1.10-636.25)
Therapies recieved	
Anti-TB	0.14 (0.03-0.47)
Antifungals	1.06 (0.30-3.46)
Antibacterials	1.28 (0.18-14.30)
Antimalarials	0.32 (0.00-16.67)
IV Fluid Recieved (per L)	0.65 (0.35-1.17)

Note:
FAMD = Factor Analysis of Mixed Data, BSI = Bloodstream infection, TB = Tuberculosis.

Table 4.10: Coefficient values (expressed as odds ratios) for models predicting death by 28 days that are linear and polynomial in time to antibacterials

	Odds Ratio	
Coefficient	Linear Model	Polynomial Model
FAMD Dimension 1	1.84 (1.45-2.38)	1.84 (1.44-2.38)
FAMD Dimension 2	1.20 (0.92-1.57)	1.19 (0.91-1.55)
Time to antibacterials (hrs)	1.01 (0.99-1.04)	1.00 (0.95-1.05)
(Time to antibacterials)^2	-	1.00 (0.99-1.02)

Note:
FAMD = Factor Analysis of Mixed Data.