Third-generation cephalosporin resistance in Blantyre, Malawi: transmission and outcomes

Thesis submitted in accordance with the requirements of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy by Rebecca Lester

June 2021

For Isaac

Collaborators

This thesis is the result of my own work. Where work was done in collaboration with others, I describe their roles here and in the relevant chapters.

Enrolment and follow-up of study participants: Research nurses Helen Mangochi, Rose Nkhata, Grace Mwaminawa and Edwin Bullah. Field workers Pilirani Kachulu, Suzgo Mkandawire and Gladys Namacha. Clinicians James Mango and Oliver Pearse.

Laboratory processing of stool, rectal swab, urine and blood culture samples: Winnie Bakali.

Laboratory processing of blood samples: MLW Blood Sciences Department.

Diagnostic bench microbiology of blood cultures: MLW Diagnostic Microbiology Department.

Building and maintaining study databases and electronic data capture forms: Lumbani Makhaza and MLW Data Department.

Statistical support in relation to study design: Brian Faragher and Chris Jewell.

The published systematic of review of 3GC-R gut mucosal carriage included as part of Chapter 2, was first-authored by Joe Lewis. I was second reviewer.

The health economic analysis for Chapter 5 was carried out by Hendramoorthy Maheswaran.

The R code to generate the household sampling points described in Chapter 2 and the R code for the transmission models in Chapter 7, was written in collaboration with Chris Jewell.

Abstract

Antimicrobial resistance occurs when microorganisms evolve to survive exposure to the antimicrobials previously successfully used to treat them. Ceftriaxone is a third-generation cephalosporin (3GC), which has long been the antibiotic of choice in many sub-Saharan African hospitals. Surveillance data from patients at Queen Elizabeth Central Hospital (QECH), in Blantyre, Malawi has shown a rapid proliferation of 3CG resistance (3GC-R) amongst key bloodstream isolates and the lack of availability of alternatives to ceftriaxone, means that these infections are frequently untreatable. Despite this, outcomes for patients with these infections in Malawi are unknown. I hypothesise that 3GC-R BSI is associated with poor outcomes for patients in Malawi, carrying a high mortality and morbidity for individual patients and a significant economic burden on the healthcare provider. Gut mucosal carriage of 3GCR-E generally precedes invasive infection and I further hypothesise that community household level determinants are driving a high prevalence of 3GC-R carriage amongst individuals living in urban Blantyre.

To address these hypotheses, I present the findings of two longitudinal cohort studies. The first, was a cohort of patients whose blood cultures were positive for Enterobacterales or *Acinetobacter spp.*. I use logistic regression and Cox proportional hazards models to determine the associations of 3GC-R on in-hospital mortality, hospital length of stay and survival. Healthcare resource use was obtained from review of the medical records and patients were interviewed to establish direct and indirect costs of admission as well as health-related quality of life (HRQoL) outcomes. I

use multivariable models to estimate the effects of 3CG-R on these health economic outcomes. The second cohort was a community sample of randomly selected households in Blantyre. Stool samples were collected from adults and children from 110 households, over 6-months and processed for 3GC-R *E. coli* using selective Chromogenic agar. I first use hierarchical models to identify risk-factors for 3GC-R gut mucosal colonisation and then develop dynamical transmission models to explore transmission routes in more detail.

I find that patients with Enterobacterales and *Acinetobacter* spp. BSI have a high mortality and that there is a significant association between 3GC-R and death, as well as increased hospital length of stay. In addition, 3GC-R is associated with higher healthcare provider and patient level costs than sensitive infection as well as poorer HRQoL outcomes. The prevalence of 3GC-R *E. coli* colonisation in the community sample is high. Sampling during rainy season and higher prevalence within households are associated with 3CG-R carriage, suggesting that within household and environmental transmission are important. Dynamical modelling provides further insight into these transmission routes, suggesting that within household reservoirs and person-to-person transmission are important drivers of 3GC-R acquisition. Future iterations of the dynamical models I develop, should incorporate social network information and whole genome sequencing of cultured isolates, in order to identify where 3GC-R transmission pathways can be interrupted.

Acknowledgements

This work would not have been possible without the patients and guardians who agreed to participate. Many did so in the face of their own critical illness, or that of their children and many did not survive. I am immensely grateful to them.

I was very lucky to work alongside a fabulous study team. They worked incredibly hard and always with good humour: Gladys Mwimaniwa, Rose Nkhata, Edwin Bullah, Winnie Bakali, Rachel Banda, Jacqueline Phiri, Gladys Namacha, Pilirani Kachulu, Suzgo Mkandawire, James Mango, Oliver Pearse and Monica Matola. In particular, I thank Helen Mangochi for her guidance and friendship.

I am grateful to my many wonderful colleagues in the UK and Malawi, for their professional and personal support, and good company: Joe Lewis (for his unwavering willingness to discuss R), Todd Swarthout, Jen Cornick, Brigitte Denis and Cat Wilson.

I thank my supervisors Nick Feasey and Chris Jewell for their tireless guidance, expertise and friendship and David Lalloo for his support and wisdom.

I am enormously grateful to the friends who supported me through this period. In particular, I am grateful to Nadja van Ginneken and Lance Turtle, who looked after me and Isaac during my write-up period. Without their kindness, this thesis would not exist.

Publications

Peer reviewed journal articles related to or arising from thesis

Lester R, Haigh K, Wood A, MacPherson EE, Maheswaran H, Bogue P, Hanger S, Kalizang'oma A, Srirathan V, Kulapani D, Mallewa J, Nyirenda M, Jewell CP, Heyderman R, Gordon M, Lalloo DG, Tolhurst R, Feasey NA. Sustained Reduction in Third-generation Cephalosporin Usage in Adult Inpatients Following Introduction of an Antimicrobial Stewardship Program in a Large, Urban Hospital in Malawi. **Clin Infect Dis**. 2020 Dec 3;71(9):e478-e486. PMID: 32060523

Lewis JM, Lester R, Garner P, Feasey NA. Gut mucosal colonisation with extended-spectrum beta-lactamase producing Enterobacteriaceae in sub-Saharan Africa: a systematic review and meta-analysis. **Wellcome Open Res**. 2019 Oct 23;4:160. PMID: 31976380

Lester R, Maheswaran H, Jewell CP, Lalloo DG, Feasey NA. Estimating the burden of antimicrobial resistance in Malawi: protocol for a prospective observational study of the morbidity, mortality and economic cost of third-generation cephalosporin resistant bloodstream infection. **Wellcome Open Res**. 2020 Jun 1;5:29. PMID: 32566760

Lester R, Musicha P, van Ginneken N, Dramowski A, Hamer DH, Garner P, Feasey NA. Prevalence and outcome of bloodstream infections due to third-generation cephalosporin-resistant Enterobacteriaceae in sub-Saharan Africa: a systematic review. **J Antimicrob Chemother**. 2020 Mar 1;75(3):492-507. PMID: 31742611

Haigh K, Dube Q, Kasambara W, Feasey NA, Lester R. Cephalosporin resistance in Malawi. **Lancet Infect Dis**. 2020 Mar;20(3):285-286. doi: 10.1016/S1473-3099(20)30047-5.PMID: 32112760

Lewis JM, Lester R, Mphasa M, Banda R, Edwards T, Thomson NR, Feasey N. Emergence of carbapenemase-producing Enterobacteriaceae in Malawi. **J Glob Antimicrob Resist**. 2020 Mar;20:225-227. PMID: 31899349

Abbreviations

3GC: Third-generation cephalosporin

3GC-R: Third-generation cephalosporin resistant/resistance

3GC-S: Third-generation cephalosporin sensitive

3GCR-E: Third-generation cephalosporin resistant Enterobacterales

3TC: Lamivudine

A&E: Accident and emergency

AETC: Adult emergency and trauma centre

APACHE: Acute physiology and chronic health evaluation

API: Analytical profile index

AMR: Antimicrobial resistance

ANTIDOTE: Antimicrobial resistance study to determine outcomes and transmission of ESBLs

aOR: Adjusted odds ratio

API: Analytical profile index

ART: Antiretroviral therapy

AST: Antimicrobial sensitivity testing

BP: Blood pressure

BSAC: British Society for Antimicrobial Chemotherapy

BSI: Bloodstream infection

CAI: Community acquired infection

CAR: Central African Republic

CD4: Cluster of differentiation 4

CDC: Centre for Disease Control and Prevention

CDDEP: Centre for Disease Dynamics, Economics and Policy

CFR: Case fatality rate

CI: Confidence interval

CNS: Central nervous system

CoM: College of Medicine

COMREC: Malawi College of Medicine Research Ethics Committee

CPAP: Continuous positive airway pressure

CPT: Cotrimoxazole preventative therapy

CRP: C-reactive protein

CSF: Cerebrospinal fluid

CVC: Central venous catheter

DAG: Directed acyclic graph

DALYs: Disability-adjusted life years

DPB: Diastolic blood pressure

DRI(s): Drug resistant infection(s)

EARS-Net: European Antimicrobial Resistance Surveillance Network

ECDC: European Centre for Disease Control

EFV: Efavirenz

EoS: Early onset (neonatal) sepsis

EPEC: Enteropathogenic *E. coli*

ESBL: Extended-spectrum beta-lactam(ase)

ESBL-E: Extended-spectrum beta-lactamase producing Enterobacterales

ETEC: Enterotoxigenic *E. coli*

EUCAST: European Committee on Antimicrobial Susceptibility Testing

FBC: Full blood count

GBD: Global Burden of Disease

GCP: Good clinical practice

GCS: Glasgow coma scale

GDP: Gross domestic product

GLASS: Global Antimicrobial Resistance Surveillance System

GNI: Gross national income

GPS: Global positioning system

H2S: Hydrogen sulphide

HAI: Hospital acquired infection

HCAI: Healthcare-associated infection

HDI: Human development index

HDU: High dependency unit

HIV: Human immunodeficiency virus

hr: Heart rate

HR: Hazard ratio

HRQoL: Health related quality of life

ICH: International Conference on Harmonisation

IQR: Interquartile range

ITU: Intensive therapy unit

IV: Intravenous

LAM: lipoarabinomannan

LIMS: Laboratory information management system

LMIC(s): Low- and middle-income country/(countries)

LoS: Late onset (neonatal) sepsis

LOS: Length of stay

LSTM: Liverpool School of Tropical Medicine

MIC: Minimum inhibitory concentration

MJC: Mercy James Centre

MLW: Malawi-Liverpool-Wellcome clinical research programme

mmHg: Millimetres of mercury

mRDT: Malaria rapid diagnostic test

MSM: Multistate model

MTB: Mycobacterium tuberculosis

MUAC: Mid-upper arm circumference

MVLR: multivariable logistic regression

MWK: Malawian Kwacha

NEQAS: National External Quality Assessment Service

ODK: Open Data Kit

OR: Odds ratio

PCP: *Pneumocystis jiroveci* pneumonia

PCR: Polymerase chain reaction

PICU: Paediatric intensive care unit

POC: Point of care

PPIs: Proton pump inhibitors

QALYs: Quality adjusted life years

QC: Quality control

QECH: Queen Elizabeth Central Hospital

QoL: Quality of life

qSOFA: Quick SOFA

RCT: Randomised controlled trial

RR: Respiratory rate

RR: Risk ratio/ respiratory rate

SBP: Systolic blood pressure

SDGs: Sustainable Development Goals

SES: Socio-economic status

SIR: Susceptible-Infectious-Recovered

SIRS: Systemic inflammatory response syndrome

SOFA: Sequential organ failure assessment

SQL: Structured query language

SpO2: Capillary oxygen saturation

sSA: Sub-Saharan Africa

TA: Traditional authority

TB: Tuberculosis

TDF: Tenofovir

UK: United Kingdom

uLAM: Urinary lipoarabinomannan

USA: United States of America

UTI: Urinary tract infection

UVA: Universal vital assessment score

UVLR: univariable logistic regression

VAS: Visual analogue scale

WASH: water, sanitation and hygiene

WCC: White cell count

WGS: Whole genome sequencing

WHO: World health organization

WHZ: Weight-for-height Z-score

Contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Overview

Antimicrobial resistance (AMR) occurs when microorganisms evolve to survive exposure to the antimicrobials previously successfully used to treat them and drug resistant infections (DRIs), which occur when AMR organisms cause disease, have been recognised as a major global public health problem. In 2014, the O'Neill Report, a major UK government commissioned review on AMR, predicted that low- and middle-income countries (LMICs) will suffer the greatest burden of DRIs, both in terms of patient outcomes and healthcare economic costs[1].

Third-generation cephalosporin resistant Enterobacterales (3GCR-E) have been highlighted by the World Health Organization (WHO) as pathogens of critical importance[2]. Ceftriaxone is a 3rd-generation cephalosporin (3GC) which has long been the antibiotic of choice in many sub-Saharan African hospitals, its once daily dosing regimen and broad-spectrum of activity favourable in settings where diagnostic and nursing capacity are limited[3].[3] Twenty years of surveillance data from patients at Queen Elizabeth Central Hospital (QECH), in Blantyre, Malawi have shown a rapid proliferation of 3CG resistance (3GC-R) amongst bloodstream Enterobacterales, occurring contemporaneously with the roll-out of ceftriaxone in 2005[3]. The lack of availability and prohibitive expense of alternatives to ceftriaxone, means that unlike in high income countries,

these infections are locally untreatable. Data on the impact of this rise in 3GC-R on patients and health systems in Malawi, are therefore urgently needed but currently lacking.

It is the hypothesis of this thesis, that bloodstream infection with 3GC-R organisms is associated with poor outcomes for patients in Malawi: carrying a high mortality and morbidity for individual patients and a significant economic burden on the healthcare provider. Gut mucosal carriage of 3GCR-E generally precedes invasive infection and I further hypothesise that community household level determinants are driving a high prevalence of 3GC-R carriage in urban Blantyre, and that diversity in human 3GCR-E *E. coli* carriage, can be explained by household level risk factors.

In this introductory chapter, I first summarise the classification and antibiotic resistance mechanisms of Enterobacterales and then review the epidemiology and drivers of human gut mucosal of 3GCR-E carriage in community settings in sub-Saharan Africa (sSA), as well as the current approaches to understanding its transmission. In doing so, I highlight the lack of available community level data in the literature. Finally, I review what is known about the burden of bloodstream infection (BSI) with 3GCR-E in sSA, both in terms of prevalence, and health and economic outcomes, placing these data in global context where relevant.
I will argue that current predictions of the impact of AMR in sSA, are in fact severely limited by lack of robust data and that the true burden of AMR in Africa is currently unknown.

1.2 Enterobacterales

1.2.1 Classification

Enterobacterales are an order of Gram negative bacteria containing seven recently defined distinct monophyletic groups of genera and over 250 species[4]. Many members of this order, including *Escherichia coli* (*E. coli)* and *Klebsiella pneumonia*e, are pathogens of major clinical importance, causing a wide range of human disease in community and hospital settings (Table 1.1). The most common Gram-negative isolates in microbiology laboratories globally[5], Enterobacterales are capable of causing infection at multiple body sites, including the urinary tract, abdominal cavity, respiratory tract and central nervous system (CNS), but perhaps worst, they are one of the commonest pathogens implicated in BSI[6].

Table 1.1**:** Medically important Enterobacterales. Adapted from[5].

Family	Genus	Species
Enterobacteriaceae	*Citrobacter*	*freundii*
		koseri
	Enterobacter	*cloacae*
		aerogenes
		sakasakii
	Escherichia	*coli*
		alvei
	Klebsiella	*pneumoniae*
		oxytoca
		granulomatis
	Kluyvera	*cryocrescens*
		ascorbata
		georgiana
		cochleae
	Raoultella	*ornithinolytica*
		planticola
	Salmonella	*enterica*
	Shigella	*dysenteriae*
		sonneii
		lexneri
		boydii
Erwiniaceae	*Pantoea*	*agglomerans*
Hafniacea	*Hafnia*	*alvei*
Morganellaceae	*Morganella*	*morganii*
	Proteus	*mirabilis*
		vulgaris
	Providencia	*stuartii*
		rettgeri
Yersiniaceae	*Yersinia*	*pestis*
		enterocolitica
		pseudotuberculosis
	Serratia	*marcesens*

1.2.2 Beta-lactam resistance in Enterobacterales

Evolving enzymatic resistance to Enterobacterales has severely limited the armamentarium of antibiotics available to treat infections with these organisms. Historically, beta-lactam antibiotics, which include penicillins

and cephalosporins, were the major class of antibiotic used, but by the 1960s, the first plasmid mediated narrow-spectrum beta lactamases (enzymes conferring resistance to the active beta-lactam ring) had been identified[7].

The subsequent rapid global dissemination of beta-lactamase resistance, prompted the development and widespread use of extended-spectrum or third-generation cephalosporins[8]. By the 1980s, plasmid-encoded enzymes, conferring resistance to these extended-spectrum oxyimino-cephalosporins had been identified[9, 10]. These extended-spectrum beta-lactamases (ESBLs) are capable of inactivating 3GCs such as ceftriaxone, ceftazidime and cefotaxime.

AmpC enzymes, which are included in functional classifications of ESBLs[11], are chromosomally encoded enzymes which are variably induced on exposure to beta-lactam antibiotics such as cephalosporins. Unlike ESBLs, they are not inhibited by clavulanic acid and may display in-vitro sensitivity to 3GC. The organisms *Enterobacter* spp., *Serratia marcesens*, *Citrobacter* spp., and *Morganella morganii*, in particular, are characterised by inducible AmpC resistance and it is generally advisable to avoid treating these organisms with 3GCs, regardless of their in-vitro susceptibility[12, 13]. Although less common, non-enzymatic resistance mechanisms such as loss or reduction of outer membrane porins[14] and modification of drug efflux systems[15] can generate in-vitro 3GC-R in Enterobacterales and not all 3GC-R isolates are therefore ESBL producers.

Laboratory detection of ESBLs typically includes a disc-diffusion screening step, using a 3GC indicator disc, followed by confirmatory testing with double or combination discs (which test for increased susceptibility in the presence of clavulanate) or with molecular methods[8]. Many laboratories in sSA, have the ability to perform the 3GC screening test, but not to confirm ESBL presence, because of the requirement for more expensive consumables and greater technical expertise[16]. Particularly since international guidelines have lowered minimum inhibitory concentration (MIC) breakpoints for many beta-lactam antibiotics, the presence of resistance to 3GCs on screening is considered sufficient evidence for likely clinical failure of these antibiotics and, where possible, for avoidance of 3GC use in patients[17].

Although the majority of 3GC-R Enterobacterales are likely to be ESBLs[18], the term third-generation cephalosporin resistance (3GC-R) or third-generation cephalosporin resistant Enterobacterales (3GCR-E) may be a more useful functional classification, especially to clinicians. Therefore, as the clinical implications of in vitro resistance to 3GCs, regardless of ESBL activity, are likely to be the same, I will use the term 3GC-R in this thesis, referring to ESBLs, but also to AmpC producers and organisms resistant to 3GC by other mechanisms.

1.2.3 Clinical implications of 3GC-R Enterobacterales

3GCR-E have been highlighted by the global AMR agenda as pathogens of major importance on which national action plans should focus their surveillance and reporting. They are listed as critical priority pathogens by the World Health Organization[2] and will be one of the first antibiotic-

bacterium combinations to be collated in future Global Burden of Disease (GBD) estimates[19]. In part, the reasons for this relate to the rising prevalence these infections globally[20] and I will discuss the epidemiology of 3GCR-E in later sections of this chapter.

A further reason, is the limited availability of alternative antibiotics to treat infections with 3GCR-E. ESBL genes are frequently located on plasmids which additionally carry genes encoding resistance to other antimicrobials, in particular aminoglycosides and fluoroquinolones[21, 22], both of which would be considered otherwise useful alternatives for invasive 3GC-R infections. In sSA, this is of particular importance, because of widespread reliance on ceftriaxone in the empirical management of sepsis[3]. In Malawi, for example, ceftriaxone is generally the first- and last-line broad-spectrum antibiotic available for treatment of severe bacterial infection and alternative antibiotics, such as meropenem and amikacin, are not routinely available[23]. Despite classification as access antibiotics on the WHO essential medicines list[24], carbapenems are prohibitively expensive in Malawi, whilst amikacin is difficult to procure due to restricted access for preservation in the management of multidrug resistant TB[25]. Unlike in high-income settings, in Malawi, these infections are therefore generally untreatable.

1.3 Gut mucosal colonisation with 3GC-R Enterobacterales

1.3.1 Introduction

Gut mucosal colonisation with 3GCR-E is a risk factor for subsequent infection[26, 27] and healthy colonised individuals are therefore thought to be a reservoir for these organisms, contributing to onward transmission and disease[28]. Interventions to reduce invasive infections with 3GCR-E may therefore be dependent on strategies aimed at reducing faecal carriage.

The prevalence of and risk factors for gut mucosal carriage with 3GCR-E in sSA have recently been systematically reviewed[29]. I was second reviewer for this article, independently performing a literature search using search terms agreed between myself and J.Lewis. I screened all abstracts and extracted data for inclusion in the review. I reviewed and edited the manuscript which was primarily written by J.Lewis, who also made the figures and tables. In this section I include data from this published paper.

1.3.2 Search strategy

The search terms and inclusion criteria used in the systematic review are shown in the appendix to this chapter (Appendix 1.9.1). The review included all studies that had screened for gut mucosal colonisation of 3GCR-E in any population in sSA, up to 18th December 2018. Here, I update the search to include papers published up to 31st December 2019 and restrict my discussion and analysis to studies which relate to colonisation in community populations, as this pertains more to my thesis

questions and hypotheses. To place the African studies in context, I additionally include a brief narrative review of studies reporting prevalence of and risk factors for 3GCR-E colonisation in other countries.

1.3.3 3GCR-E colonisation in sSA: prevalence and risk factors

Table 1.2 shows 3GCR-E carriage prevalence estimates from the 12 identified community-based studies providing data from sSA. Prevalence is described as proportions of 3GC-R calculated from numbers of isolates of *E. coli* and *Klebsiella spp.*, tested against a 3GC in each study. A forest plot was generated, illustrating proportion estimates for each study with 95% confidence intervals (95% CI), calculated using the Wilson's score method. The I^2 statistic was calculated to quantify heterogeneity and also computed for subgroup differences.

A forest plot of prevalence estimates for all included studies is shown in Figure 1.1 Prevalence ranged from 5.0-59.0% (median 16.0%). Overall heterogeneity of prevalence estimates was high (I^2=95%) and not explained by subgroup analysis by age group of population (test for subgroup differences, p=0.35) (Figure 1.1). The notable difference in these studies, which may explain some of this heterogeneity, is in the study populations, which include nomadic pastoralists, pregnant women, street children and university students. Only one study was a random sample of the general population, so although African 3GCR-E carriage rates appear high, unselected community population prevalence is really unknown.

Table 1.2: Studies from sSA reporting resistance to 3GC in gut mucosal Enterobacterales in community populations. Adapted from[29].

First author, year	Country	Study type	Inclusion population	n	Age group	Model	Prevalence n/N(%)	Risk factors assessed	Significant risk factors
Albretchova, 2012[30]	Kenya	Cross-sectional	Nomadic pastoralists	23	Adults	ND	4/23(17.4)	ND	NA
Chereau, 2015[31]	Madagascar	Cross-sectional	Pregnant women	356	Adults	MVLR	66/356(18.5)	Age, study area, marital status, education, electricity access, type of birth attendant, housing type, toilets, water, animals at home, hospitalisation, abx use	Private access to inside drinking water
Chirindze, 2018[32]	Mozambique	Cross-sectional	University students	275	Adults	ND	35/275(12.7)	ND	NA
Farra, 2016[33]	CAR	Cross-sectional	Healthy controls in diarrhoeal study	134	Children	MVLR	79/134(59.0)	Age, gender, comorbidity, SES, nutritional status, animals at home, toilets, urban/rural, no. hh members, eating habits, water	Highest vs lowest SES
Katakwebe, 2018[34]	Tanzania	Cross-sectional	General population	70	Adults	ND	21/70(30.0)	ND	NA
Lonchel, 2012[35]	Cameroon	Cross-sectional	Student volunteers	150	Adults	UVLR	10/150(6.7)	Age, male gender, antibiotic use	None found
Mahamat, 2019[36]	Chad	Cross-sectional	Student or healthcare worker volunteers	100	Adults	ND	29/100(29.0)	ND	NA

Table 1.2: Studies from sSA reporting resistance to 3GC in gut mucosal Enterobacterales in community populations. Adapted from[29] (continued).

First author, year	Country	Study type	Inclusion population	n	Age group	Model	Prevalence n/N(%)	Risk factors assessed	Significant risk factors
Moremi, 2017[37]	Tanzania	Cross-sectional	Street children	107	Children	MVLR	34/107(31.8)	Age, education, herb use, source of income, source of food, street child type	Local herb use, sleeping on streets vs not
Ribeiro, 2016[38]	Angola	Cross-sectional	No antibiotics/hospital exposure prior 3months	18	Adults	ND	4/18(22.2)	ND	NA
Ruppe, 2009[39]	Sengal	Cross-sectional	Remote villages, CTX-M ESBl only	20	Children	ND	2/20(10.0)	ND	NA
Sanneh, 2018[40]	Gambia	Cross-sectional	Food handlers in schools	565	Adults	UVLR	28/565(5.0)	WASH behaviours, hospitalisation within 3m, abx from street, completing abx, diarrhoea/UTI in prior 3m, food handling training	Lack of food handling training, abx within 3m
Tellevic, 2016[41]	Tanzania	Cross-sectional	<2yr attending health centre for vaccine	250	Children	MVLR	29/250(11.6)	Age, HIV status, gender, residence, parental education, nutritionalstatus, abx in prior 14 days	HIV infected, Kinondoni district, abx use in prior 14 days

Note:
CAR, Central African Republic, MVLR = Multivariable logistic regression, NA = Not applicable, ND = Not done, SES = Socio-economic status, UTI, Urinary tract infection, UVLR = Univariable logistic regression, abx = antibiotics, hh = household, m=months

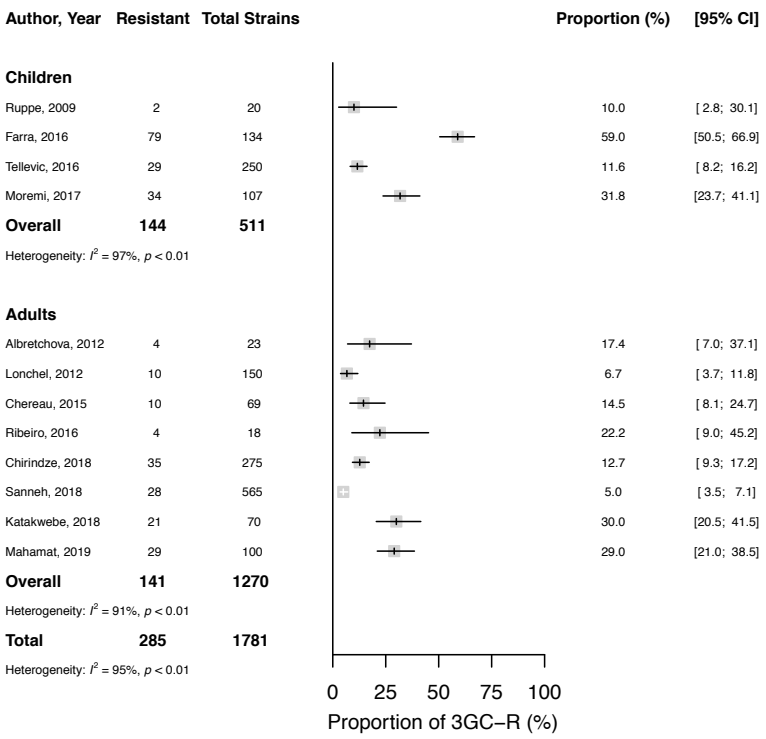


Figure 1.1: Forest plot of prevalence of 3GC-R by study, stratified by age-group of population. Test for subgroup differences: I2 = 0.87, df = 1, p=0.35. Adapted from[29].

Five studies in sSA explored risk factors for 3GCR-E in community populations (Table 1.2). Prior antibiotic use in the preceding two weeks or three months, was identified as a significant risk factor in two out of four studies that explored it[40, 41]. The role of water, sanitation and hygiene (WASH) practices is difficult to elucidate, having only been explored as risk factors in three studies and found to be significant in one[31]. This study unexpectedly found that private access to indoor drinking water was significantly associated with carriage, which is inconsistent with suggested faeco-oral transmission routes for 3GCR-E, but may relate to increased access to healthcare and antibiotics associated with higher socio-economic

status (SES)[31]. Higher SES was associated with 3GCR-E carriage in a study from the Central African Republic, but again, access to healthcare and cost of antibiotics was not discussed. Finally, the data relating HIV to carriage risk are surprisingly sparse. Only one study included HIV status as a covariate in their regression model, finding it to be a significant risk factor for 3GCR-E carriage in Tanzanian children attending a health centre[41].

Although not included in Table 1.2, which describes only community level studies, hospitalisation appears to be a risk factor for carriage, as evidenced by higher prevalence estimates in inpatient cohorts than community cohorts in sSA, with prevalence estimates ranging from 45-60%[29].

There are currently no published reports of 3GC-R carriage prevalence from Malawi, but unpublished data suggests a high prevalence. A cross-sectional study of 103 adult inpatients at QECH in 2009-2010, found an overall 3GCR-E prevalence of 49% and that carriage was associated with hospital length of stay and prior hospitalisation (K.Gray, unpublished). A larger and more recent (2017-2018) cohort of adult sepsis patients found a baseline 3GCR-E carriage prevalence of 49% (95% CI 32-52%) amongst sepsis patients on admission to QECH and of 41% (95% CI 32-52%) in antibiotic-unexposed inpatients, also on admission (J.Lewis, unpublished). Amongst healthy community controls recruited in this study, carriage prevalence was 28% (95% CI 20-38%), suggesting that within household transmission may be important (J.Lewis, unpublished). Prior to this thesis there have been no large scale purely community-based studies of 3GCR-E carriage from Malawi.

1.3.4 3GCR-E colonisation in global context

Human faecal carriage with 3GCR-E is present in community settings worldwide, but with significant between and within country differences in prevalence estimates[42]. Higher carriage rates are typically seen in Asia, with community prevalence of 50-80% in studies from China [43, 44], and 50% in Laos and Vietnam[45, 46]. In contrast, European prevalence estimates of are typically less than 10% with recent community based studies finding prevalence of 7.3% in the UK[47], 4.5% in the Netherlands[48] and 4.7% in Sweden[49]. The African estimates described above therefore frequently approach the high prevalence of 3GCR-E carriage found in Asia.

The risk factors for 3GCR-E carriage which are consistently identified across multiple studies, are exposure to healthcare facilities[50, 51], antibiotic use during the previous 1-4 months[51-53] and travel from low to high prevalence areas[42, 53-55]. Other studies have identified use of proton pump inhibitors (PPIs)[56, 57] and exposure to food-producing animals as risk factors[58]. As with the African literature, there is a sparsity of data evaluating household-level risk factors[42].

1.3.5 Dynamics of gut mucosal carriage with 3GCR-E: longitudinal data and approaches to understanding transmission

As described in the previous section, the majority of 3GCR-E carriage data are derived from cross-sectional studies which use a logistic regression analysis to explore risk factors. A small number of studies from outside of sSA have collected longitudinal data. A large study in the Netherlands

investigated risk factors for persistent ESBL carriage in healthy community volunteers who provided 5 monthly faecal samples over an 8-month period, finding that 32.9% of participants colonised at baseline remained positive in all follow-up samples (median duration 242 days)[48]. Antibiotic exposure, PPI use and *E. coli* phylogroup B2 or D were associated with carriage persistence. Other European studies find variable carriage duration rates, ranging from 4.3 months in a French study of persistent carriage following hospitalisation[59], to 33 months in a German study of patients discharged following an outbreak in a hospital and long-term care facility[60]. The reason for these differences is not clear, but may in-part relate to different sampling intervals and follow-up durations in these studies.

In an attempt to overcome the issues of interval-censored data experienced by studies which have sparse sampling intervals, one small study of 23 travellers to Laos, undertook daily sampling of participants for a three-week period[61]. All participants were colonised with 3GCR-E by the end of their travel, but status of individuals varied day by day. Using whole genome sequencing (WGS), the study demonstrated that the majority of participants acquired multiple strains throughout follow-up, suggesting multiple transient colonisation events. The Dutch study described above, also found multiple different *E. coli* sequence types in longitudinal samples from the same individuals[48].

Despite the lack of longitudinal data, there is a developing signal from the literature that 3GCR-E carriage is a highly dynamic process. Additionally, there is evidence that person-to-person transmission within households is

important, with three studies including one from Madagascar, finding colonisation of a household member to be a risk factor for carriage in multivariate logistic regression[62-64].

Cross-sectional data and logistic regression models cannot effectively parameterise dynamic processes involving person-to-person transmission. At most, they can incorporate number of people in a particular environment (hospital ward or household) as a covariate, but they cannot capture what is likely to be a time-varying force of infection on an uncolonised individual, from colonised individuals who may share the same environment.

Conceptually, in transmission of infectious diseases, incidence of new infection (or colonisation) depends on prevalence, i.e. new cases of infection (or colonisation) occur at a rate proportional to the number of infected (or colonised) and susceptible patients in a particular environment[65]. Dynamical transmission models attempt to accurately capture these transmission dynamics and to overcome the limitations of traditional regression models, by parameterising a time-varying force of infection on a susceptible individual from infectious individuals. These models, which are largely based on the general Susceptible-Infectious-Recovered (SIR) epidemic model[65, 66], are potentially a far more realistic representation of dynamical colonisation processes in which person-to-person transmission is likely to be important.

To date, this approach has generally been generally applied to modelling 3GC-R carriage within hospital environments. A Cambodian study fitted a

transmission model to data collected from a longitudinal carriage study in a neonatal ITU, incorporating time-varying covariates to assess within-ward transmission of 3GC-R *Klebsiella pneumoniae* and finding evidence of person to person spread within wards[67]. Only one published study has taken this approach to 3GCR-E transmission in sSA, again with a hospital-based cohort. This study, from Madagascar, fit a dynamical model to longitudinally collected newborn 3GCR-E acquisitions, finding that acquisition sources were species dependent, but that transmission from healthcare workers was important[68].

Very few community-based studies have used dynamical models and accurate data describing within household transmission of 3GCR-E are therefore scarce. Two studies from the Netherlands have taken this approach. The first, used a transmission model to fit data from hospitalised patients colonised with ESBLs and from their household members over an 18-month period, showing evidence of within household transmission[69]. The second was a population-based transmission model which fit 3GCR *E. coli* data from multiple sources (animals, food, humans, water), finding evidence of person-to-person transmission had a greater impact on transmission than direct animal contact, food consumption or environmental sources[70]. The role of within household transmission of 3GCR-E in sSA is currently unknown.

1.4 Bloodstream infection with 3GCR-E: measuring the healthcare burden

BSIs generally have the most severe clinical impact of drug-resistant infections and being easily defined in the laboratory, are considered the most useful type of infection to include in international AMR surveillance networks[18, 71].

Prevalence and mortality are key measures of healthcare burden[6], and in this section, I present here the main methods and results of my recently published systematic review of prevalence and mortality outcomes of 3GCR-E BSI in sSA[16]. The published review included *E. coli*, *Klebsiella spp.* and *Salmonella* spp. but Salmonellae are excluded here since they have not yet displayed 3GC-R in Malawi[16] and are therefore not included in any part of the study described in this thesis. I additionally include narrative review of the global literature, to place the African data in context and to further understand the epidemiology and global burden of 3GCR-E BSI.

I subsequently review three other key measures of the health burden of 3GCR-E BSI: hospital length of stay, economic cost and patient quality of life. Where possible I focus on literature from sSA, but place this in global context, particularly where African data are lacking.

1.4.1 Prevalence and mortality in sSA

1.4.1.1 Search strategy and inclusion criteria

The literature was systematically reviewed for studies published between 1st January 1990 and 31st December 2019, to identify clinical cohorts and case-controls studies reporting 3GC-R susceptibility testing of *E. coli* and *Klebsiella* spp.. The review aimed to answer the following questions. Firstly, what is the prevalence of 3GC-R amongst bloodstream infection in sSA? Secondly, what is the mortality from these infections in sSA?

Pubmed and Scopus were searched using search terms shown in the appendix to this chapter (Appendix 1.9.2). Studies were included if they tested *E. coli* or *Klebsiella spp.* for 3GCR. Methods of confirmatory ESBL testing, such as double-disc synergy or PCR, were extracted from articles if they were reported, but we did not exclude studies that did not confirm ESBL status. Surveillance data, in addition to studies reporting clinical cohorts were included, but case reports, case series, expert opinions and reviews were excluded. References cited in selected articles were reviewed for additional articles and authors were contacted to obtain original data, where percentages but not absolute numbers of resistant organisms were provided.

1.4.1.2 Statistical analysis

Prevalence is described as a proportion of 3GC-R, calculated from numbers of isolates of *E. coli*, *Klebsiella* spp., tested against a 3GC and the number of resistant strains. Forest plots were generated, illustrating proportion estimates for each study with 95% CI calculated using the Wilson's score

method. The I^2 statistic was calculated to quantify heterogeneity. High levels of heterogeneity amongst included studies precluded meaningful meta-analysis, and I therefore present median prevalence of 3GC-R for each pathogen, with corresponding IQR to provide an assessment of the wide range in resistance prevalence. Medians were calculated for sSA and for each Africa Region as defined by the United Nations Statistics Division[72].

Heterogeneity of proportion estimates was explored using pre-defined subgroup analysis by Africa region and a post-hoc subgroup analysis by age-group of study population. Visual inspection of resulting forest plots was carried out and a test for subgroup differences applied where visual inspection suggested a likely difference in subgroup proportion estimates, and where more than two studies contributed to each subgroup. Additionally, I examined for trends in proportions estimates over time using visual inspection of forest plots ordered by year of publication, and a linear meta-regression model.

1.4.1.3 Study characteristics

A total of 31 studies, comprising 8,251 isolates, reported prevalence of 3GC-R amongst non-Salmonella bloodstream Enterobacterales and these are summarised in Table 1.3. Of these, 20 studies reported proportions of 3GC resistance in *E. coli* and 28 in *Klebsiella spp.* Data were available from only 10 countries across three sSA regions and were biased towards South Africa (Figure 1.2). All studies were observational. Eleven studies were laboratory based, reviewing isolates with or without associated hospital records, including one multi-site surveillance study. Twenty

studies recruited patients: 4 cross-sectional, 15 cohort studies (10 prospective, four retrospective, one mixed) and one case-control study. Many studies were in paediatric populations (17/31), including six exclusively in neonates. Three studies recruited adults over 16 years of age, 11 recruited in all age groups and one study did not report age of participants.

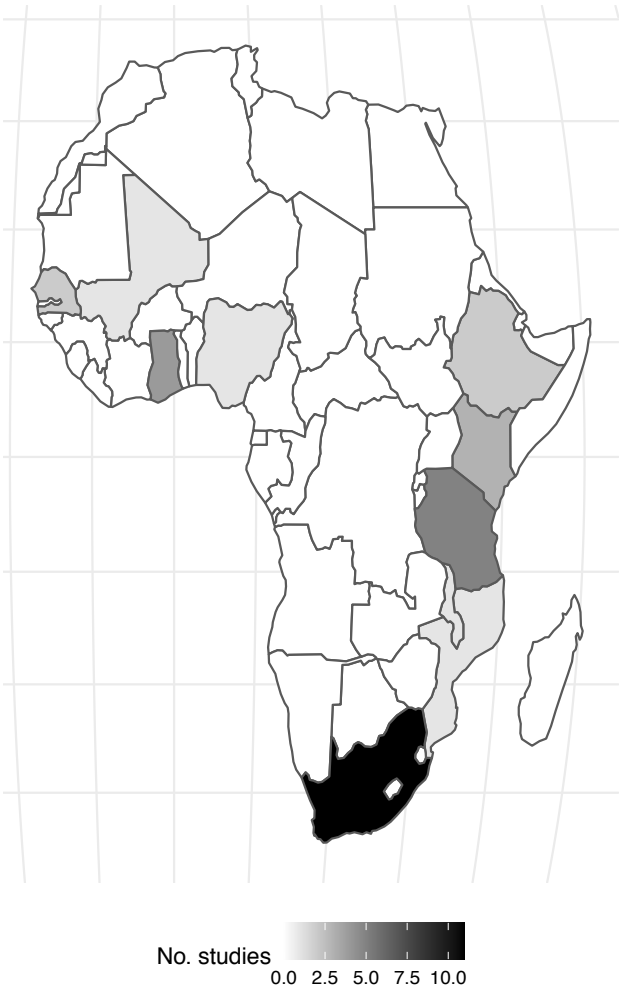


Figure 1.2: Geographical location of studies reporting 3GC susceptibility testing of bloodstream Enterobacterales. Adapted from[16].

Table 1.3: Studies from sSA reporting resistance to 3GC in bloodstream Enterobacterales. Adapted from[16].

Year	First author	Country	Patient Population	Laboratory methods Culture AST	HIV prevalence n/N (%)	Blood culture positivity n/N (%)	HAI/CAI	*E. coli* n/N (%)	*Klebsiella* spp. n/N (%)
2013	Acquah[73]	Ghana	IP U RH P	A DD	NR	86/331 (26.0)	NR	ND	1/12 (8.3)
2016	Apondi[74]	Kenya	IP U RH All	A DD	NR	NR	NR	ND	68/78 (87.2)
2005	Bejon[75]	Kenya	IP R DH P	M+A Etest	NR	NR	NR	0/141	4/63 (6.0)
2007	Blomberg[76]	Tanzania	IP U RH P	A DD+Etest	16.8%	255/1828 (13.9)	HAI+CAI	9/37(24.3)	9/52 (17.0)
2016	Breurec[77]	Senegal	IP U RH N	M DD	NR	77/226 (34.0)	EoS+LoS	ND	33/39 (84.6)
2007	Brink[78]	South Africa	IP U RH All	NR DD+V	NR	NR	NR	47/471 (10.0)	293/636 (46.0)
2016	Buys[79]	South Africa	IP U RH P	A DD+V+Etest	82/410 (20.0)	NR	HAI+CAI	ND	339/410 (83.0)
2018	Crichton[80]	South Africa	IP U RH P	A DD+V	18/141 (12.8)	938/7427 (12.6)	HAI+CAI	8/36 (22.0)	9/9 (100)
2015a	Dramowski[81]	South Africa	IP U RH N	A V	NR	717/6251 (11.5)	All HAI	7/58 (12.1)	172/235 (73.2)

Table 1.3: from sSA reporting resistance to 3GC in bloodstream Enterobacterales. Adapted from[16] (continued).

Year	First author	Country	Patient Population	Laboratory methods Culture AST	HIV prevalence n/N (%)	Blood culture positivity n/N (%)	HAI/CAI	*E. coli* n/N (%)	*Klebsiella* spp. n/N (%)
2015b	Dramowski[82]	South Africa	IP U RH P	A V	13.4%	935/17001 (5.5)	HAI+CAI	12/97 (12.4)	122/158 (77.2)
2016	Eibach[83]	Ghana	IP U RH All	A V	NR	NR	HAI+CAI	5/50 (10.0)	34/41 (82.9)
2008	Jaspan[84]	South Africa	IP U RH P	NR DD+Etest	100% (HIV cohort)	NR	HAI (*Klebsiella*)	ND	11/11(100)
2002	Ko[85]	South Africa	IP U RH Ad	N	7/40 (18.0)	NA	CAI	ND	3/40 (7.5)
2012	Kohli[86]	Kenya	IP U RH All	A DD	123/1092 (11.3)	1092/18750 (5.8)	NR	10/69 (14.5)	5/38 (13.1)
2017	Lochan[87]	South Africa	IP U RH P	A V+DD+Etest	17/524 (13.4)	958/16,951 (5.7)	CAI,HAI and HCAI	31/92 (33.7)	68/88 (77.3)
2018	Marando[88]	Tanzania	IP R DH N	M DD	NR	60/304 (19.7)	NR	ND	21/26 (80.1)
2012	Mhada[89]	Tanzania	IP U RH N	M DD	NR	5/330 (1.5)	NR	2/14 (14.3)	4/22 (18.2)
2014	Morkel[90]	South Africa	IP U RH N	NR	HIV exposed 9/54 (16.6)	58/503 (11.5)	NR	ND	10/17 (58.8)
2009	Mshana[91]	Tanzania	IP U RH NR	NR	NR	NR	NR	ND	29/31 (93.5)

Table 1.3: from sSA reporting resistance to 3GC in bloodstream Enterobacterales. Adapted from[16] (continued).

Year	First author	Country	Patient Population	Laboratory methods Culture AST	HIV prevalence n/N (%)	Blood culture positivity n/N (%)	HAI/CAI	*E. coli* n/N (%)	*Klebsiella* spp. n/N (%)
2017	Musicha[3]	Malawi	IP U RH All	A DD	NR	29,183/ 194,539 (15)	NR	140/1311 (10.7)	260/542 (48.0)
2016	Ndir[92]	Senegal	IP U RH P	NR	NR	173/1800 (9.6)	NR	7/12 (58.3)	33/ 40 (82.5)
2013	Obeng-Nkrumah[93, 94]	Ghana	IP U RH All	A DD	NR	NR	NR	5/17 (29.4)	13/26 (50.0)
2016	Obeng-Nkrumah[93]	Ghana	IP U RH P	A DD	NR	1451/15683 (9.3)	NR	63/112 (56.2)	40/68 (58.8)
2011	Ogunlesi[95]	Nigeria	IP U RH N	M DD	NR	174/1050 (16.6)	NR	6/16 (37.5)	12/33 (36.4)
2015	Onken[96]	Tanzania (Zanzibar)	IP U RH All	M+A DD+V	NR	66/470 (14.0)	NR	1/10 (10)	5/11 (45.5)
2004	Paterson[97]	South Africa	IP U RH Ad	M+A NR	NR	NR	NR	ND	28/76 (37.0)
2014	Perovic[98]	South Africa	IP U RH All	NR Microscan	NR	NR	NR	ND	1895/2774 (68.3)
2015	Preziosi[99]	Mozambique	IP U RH Ad	A DD	652/841 (77.5)	63/841 (7.5)	NR	1/14 (7.1)	ND
2016	Sangare[100]	Mali	IP U RH All	A DD	NR	NR	NR	8/34 (23.6)	10/34 (29.4)

Table 1.3: from sSA reporting resistance to 3GC in bloodstream Enterobacterales. Adapted from[16] (continued).

Year	First author	Country	Patient Population	Laboratory methods	HIV prevalence	Blood culture positivity	HAI/CAI	*E. coli* n/N (%)	*Klebsiella* spp.
2015	Seboxa[101]	Ethiopia	IP U RH All	A+M DD	123/399 (30.1)	38/299 (12.7)	NR	8/16 (50)	30/35 (85.7)
2015	Wasihun[102]	Ethiopia	IP U RH All	A DD	NR	NR	NR	9/16 (56.2)	ND

Note:
AST = Antimicrobial susceptibility testing, BSI = Bloodstream infection
IP = Inpatient, U = Urban, R = Rural, RH = Referral hospital, DH = District hospital
Ad = adults, P = paediatric, N = neonates, All = all ages
A = automated, M = manual, DD = Disc diffusion, V = Vitek
CAI= Community acquired infection, HAI = Hospital acquired infection, HCAI = Healthcare associated infection, EoS = Early onset sepsis, LoS = Late onset sepsis,

1.4.1.4 Prevalence estimates

Median estimates of 3GC resistance in *E. coli*, *Klebsiella* spp. are shown in Table 1.4, together with median estimates by Africa region and forest plots of individual studies are shown in Figures 1.3 and 1.4. The median point estimate of 3GC-R in *E. coli* BSI from 20 studies, was 18.4% [IQR 10.5 to 35.2] (Table 1.4). Heterogeneity was high (I^2=93%) (Figure 1.3) and not explained by subgroup analysis by Africa region (Appendix Figure 1.1). Median point estimates of 3GC resistance in *Klebsiella* BSI were higher across all regions than for *E. coli* with an overall estimate of 54.4% [IQR 24.3 to 81.2] from 28 studies (Table 1.4, Figure 1.3). As with *E. coli*, heterogeneity was high (I^2=96%) and not explained by differences in Africa region (Appendix Figure 1.1).

Table 1.4: Median prevalence of 3GC-R resistance in E coli and *Klebsiella* spp., BSI, shown by Africa region. Adapted from[16].

Pathogen	Overall 3GC-R Prevalence % [IQR]	Eastern Prevalence % [IQR]	Middle Prevalence % [IQR]	Western Prevalence % [IQR]	Southern Prevalence % [IQR]
E. coli	18.4 [10.5 to 35.2] 20 studies	14.3 [10.0 to 24.3] 9 studies	No data	33.5 [25.0 to 51.6] 6 studies	12.4 [12.1 to 22.2] 5 studies
Klebsiella spp.	54.4 [24.3 to 81.2] 28 studies	46.7 [17.3 to 84.5] 10 studies	No data	58.3 [IQR 34.6 to 82.6] 8 studies	63.6 [39.1 to 76.2] 10 studies

Note:
IQR=Interquartile range

The heterogeneity in prevalence estimates was not explained by subgroup analysis by Africa region, nor by age group of study population (Appendix Figure 1.2), but instead may reflect the diverse populations sampled and

the variety of laboratory microbiological methods used, both for organism identification and for antimicrobial susceptibility testing (Table 1.3). Prevalence of resistance amongst key pathogens is also likely to be influenced by a variety of clinical parameters including HIV status, healthcare attendance and prior antibiotic use, but these data were rarely reported (Table 1.3) and subgroup analysis by these factors was therefore impossible.

The classification of isolates by source, for example whether community-acquired infection (CAI), hospital-acquired infection (HAI), or healthcare-associated infection (HCAI) is key to interpretation of these data. Twenty-nine studies tested bloodstream isolates from patients presenting to public referral or private hospitals in urban settings, with only two recruiting from rural district hospitals (Table 1.3). Six studies investigated the difference in blood culture pathogens and prevalence of resistance between CAI and hospital acquired HAI or HCAI. Of these, five found a higher prevalence of 3GC-R in HAI. Two studies were cohorts of patients with hospital acquired infection and one study included only patients with suspected community acquired BSI. Of the six neonatal studies, two differentiated early-onset from late-onset neonatal sepsis but did not report on differences in proportions of 3GC resistance between the two groups.

Long term surveillance data from sSA are scare, and interpreting the significance of proportion estimates in the absence of trend data is challenging. The first studies reporting 3GC-R in *Klebsiella* spp. are from 2004 and in *E. coli* from 2005 (Table 1.3). Metaregression incorporating year of study publication as a fixed effect, demonstrated a clear increase in

prevalence of 3CG-R for *Klebsiella* spp. over time (Figure 1.3) and a non-significant increase in prevalence of 3GC-R *E. coli* (Figure 1.4). The most comprehensive published prevalence data comes from Malawi, where 20 years of blood culture surveillance from patients of all ages presenting to Queen Elizabeth Central Hospital. These data revealed a decline in the overall incidence of BSI, in parallel to public health interventions such as ART roll out and malaria control, but a rapid expansion of in 3GC-R amongst Enterobacterales, in particular *E. coli* and *Klebsiella* spp. since 2005, in parallel to roll-out of ceftriaxone as the intravenous antibiotic of choice for treatment of sepsis and severe bacterial infection in sSA[3].

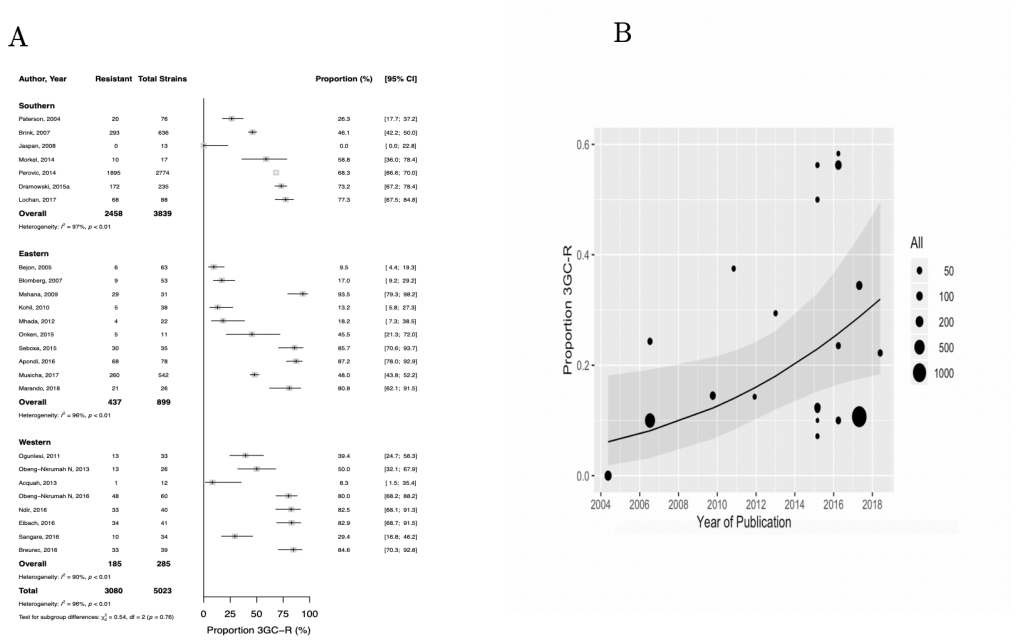


Figure 1.3: 3GC-R *Klebsiella* spp. in bloodstream infection in sSA. A: Forest plot of proportions of 3GC-Rstudy and by Africa region. Test for subgroup differences: I2 = 3.09, df = 2, p=0.21 (random effects model) B: Logistic mixed effects regression of proportion estimates as a function of year of publication (significant association with proportion of 3GC-R, p<0.01). Adapted from[16].

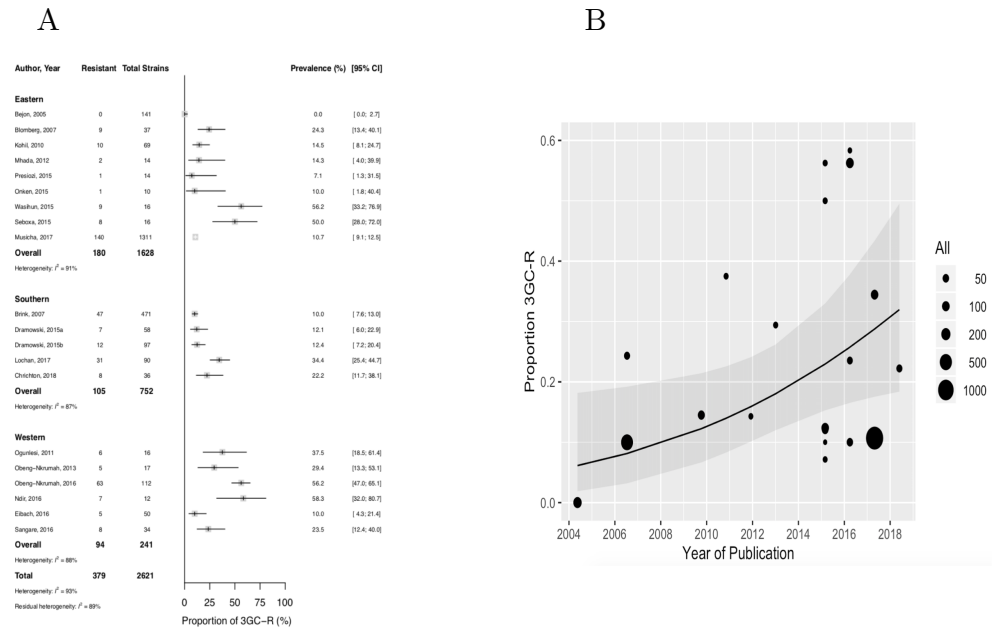


Figure 1.4: 3GC-R *E. coli* in bloodstream infection in sSA. A: Forest plot of proportions of 3GC-R by study and by Africa region. Test for subgroup differences: I2 = 0.54, df = 2, p = 0.76 (randomeffects model) B: Logistic mixed effects regression of proportion estimates as a function of year of publication (non-significant association with proportion of 3GC-R, p=0.02). Adapted from[16].

1.4.1.5 Mortality burden from 3GCR-E in sSA

The Malawian dataset, though comprehensive in its prevalence estimates, does not link to clinical metadata, other than patient age and sex[3]. Literature review identified only two studies from sSA (both from Senegal) that were designed specifically to determine attributable mortality from 3GC-R (Table 1.5). The first, a case-control study in children recruited 110 cases and 220 controls, finding that 3GC-R was the only significant independent risk factor for death in a multivariable logistic regression model, adjusting for age, admission diagnosis, and inadequate initial antibiotic therapy (OR 2.9, 95% CI 1.8-7.3)[92]. The second, by the same investigators, was a retrospective cohort in children. Case-fatality rate (CFR) was higher in 3GC-R infection (47.3%) than in 3GCR-S infection (22.4%) with a significant impact of 3GC-R on multivariable analysis (OR 5.3, 95% CI 3.1-19.9).

A further six studies reported a mortality statistic from 3GC-R BSI in sSA (Table 1.5), but these were not designed to determine attributable mortality. Rather, these were sepsis or bloodstream infection cohorts, aiming to identify risk factors for death in sepsis, and either including 3GC-R as an explanatory variable in their models or more simply reporting a crude case fatality rate. Most of these studies were recruiting in paediatric populations, but the mixture of retrospective and prospective designs, and heterogenous outcome measures (odds ratios, relative risks, or case-fatality rates) prohibit meta-analysis and estimation of a pooled mortality estimate from 3GC-R in sSA.

Where available, case-fatality rates from 3GC-R BSI in individual studies appear to be high, ranging from 17-100%. Two studies concluded that 3GC-R was an independent risk factor for mortality, though in one prospective cohort recruiting patients of all ages, this effect was only seen when neonates were removed from the analysis. In a Tanzanian prospective cohort study, the significant effect of 3GC-R on mortality appeared to be mediated through inappropriate empirical antibiotics, and 3GC-R alone was not an independent risk factor for death. There were no survival analysis models identified in the literature from sSA.

Despite the rising prevalence of 3GC-R in invasive organisms, and a reliance on ceftriaxone for empirical management of sepsis, the clinical impact of 3GC-R BSI in sSA has therefore not yet been described.

Table 1.5: Studies from sSA reporting mortality from 3GCR-E. Adapted from[16].

First author, year	Country	Study type	Population	Pathogen	Inclusion criteria	Sample size	Outcome	Outcome period	Model	Author conclusions
Ndir, 2016[103]	Senegal	Retrospective cohort	Paediatric	Mixed	Clinical signs of infection	3GC-R 110 3GC-S 76	CFR (3GC-R) 47.3% CFR (3GC-S) 22.4% OR 5.3(3.1-19.9)	In-hospital	Multivariable; attributable mortality from 3GC-R BSI	3GC-R is associated with fatal outcome
Onken, 2015[96]	Tanzania (Zanzibar)	Prospective cohort	All ages	Mixed	SIRS criteria or suspected systemic bacterial infection	3GC-R 7 3GC-S 22	CFR (3GC-R) 60.0% CFR (3GC-S) 36.0%	In-hospital	Not done	No significant difference in CFR between 3GC-R/3GC-S but small numbers

Table 1.5: Studies from sSA reporting mortality from 3GCR-E. Adapted from[16] (continued).

First author, year	Country	Study type	Population*	Pathogen**	Inclusion criteria	Sample size	Outcome	Outcome period	Model	Author conclusions
Seboxa, 2015[101]	Ethiopia	Prospective cohort	Adults	Mixed	SIRS and suspected sepsis	3GC-R 11 3GC-S 9	CFR (3GC-R) 100% CFR (3GC-R) 11.0% RR: 9.00(1.4-57.1)	In-hospital	Not done	3GC-R associated with fatal outcome
Buys, 2016[79]	South Africa	Cross-sectional	Paediatric (including neonates)	*Klebsiella* pneumoniae	*Klebsiella* isolates from electronic database	3GC-R 339 3CG-S 71	CFR NR OR 1.09(0.55 – 2.66)	In-hospital	Multivariate; risk factors for mortality from *Klebsiella* BSI	Incorrectly conclude an association of 3GC-R with mortality
Eibach, 2016[83]	Ghana	Prospective cohort	All ages	Mixed	Fever or history of fever (all ages) and neonatal sepsis	3GC-R 41 3GC-S 377	OR 3.0(1.2-7.3)all 0.6(0.1-3.7)neo	In-hospital	NR	3GC-R BSI is associated with higher mortality, dependent on age.

Table 1.5: Studies from sSA reporting mortality from 3GCR-E. Adapted from[16] (continued).

First author, year	Country	Study type	Population	Pathogen	Inclusion criteria	Sample size	Outcome	Outcome period	Model	Author conclusions
Crichton, 2018[80]	South Africa	Prospective cohort	Paediatric (including neonates)	Mixed	Clinical signs or risk factors for severe bacterial infection	3GC-R 13 3GC-S 38	CFR (3GC-R) 23.1%	In-hospital	Multivariate; risk factors for mortality from sepsis (3GC-R not included in model)	Crude mortality higher in 3GC-R than in other BSI

Note:
*Adult definition included adolescents (aged 13 years and above)
**Mixed refers to mixed Enterobacterales not differentiated by species in the analysis
CFR = Case-fatality rate, OR = Odds ratio, RR = Risk ratio, SIRS = Systemic inflammatory response syndrome, neo=neonates

1.4.2 3GCR-E BSI in global context: prevalence and mortality

The European Antimicrobial Resistance Surveillance Network (EARS-Net) is the largest international surveillance dataset available, reporting AMR data from over 1400 European hospitals via laboratory surveillance[18]. These data have demonstrated a gradual rise in 3GC-R amongst bloodstream *E. coli* since 2000, with overall prevalence reported at 10-12% in 2018[18]. There is significant between country heterogeneity, with estimates generally higher in South-eastern Europe than in other regions (28.7% in Italy, 19.3% in Greece, 11.0% in the UK, 7.7% in Denmark). The Centre for Disease Dynamics, Economics and Policy (CDDEP), collate national AMR data from countries worldwide, though these data are derived from multiple disparate sources including private hospitals and laboratories and may be less reliable[104]. CDDEP reports suggest that prevalence of 3GCR in BSIs appears to be much higher in middle-income countries at 77.0% in India, 64.0% in China and 71.0% in Vietnam (2017)[104, 105] .

As with the African estimates, 3GC-R prevalence amongst *Klebsiella* spp., is invariably higher than in *E. coli*, at 27% overall in Europe[18]. Prevalence is again higher in middle-income countries, at 77% in India, 55% and 51% in Vietnam[104]. The reasons for these country and regional level differences are not clear and are difficult to elucidate at country level, but may relate to differences in antimicrobial consumption[106] and economic status[105].

Representative studies investigating mortality from 3GCR-E are summarised in Appendix Table 1.1. Large studies from high-income

countries generally find association between 3GC-R and mortality, regardless of design, comparator group or outcome measure. The BURDEN study remains the largest prospective study of attributable mortality from 3GCR BSI to date[107]. Using a matched parallel cohort design, the study recruited just over 3500 patients in 13 European countries, comparing 3GC-R infections to two control groups: patients with 3GC-S BSI and non-infected patients, the latter matched for hospital length of stay before enrolment. Patients with 3GC-R BSI had a crude mortality of 36% and a higher in-hospital mortality in a logistic regression model when compared to non-infected patients (OR 4.6, 95%CI 1.7-12.3) and when compared to 3GC-S BSI (OR 1.9, 95% CI 1.4-2.4). Cox regression models confirmed the effect of 3GC-R on mortality, compared to 3GC-S BSI (HR 2.9, 95%CI 1.2–6.9) and compared to non-infected controls (HR 5.7 95%CI, 2.5–13.0)[107].

Two other studies used a similar matched parallel cohort design. The first, recruited over 600,000 European acute care admissions in a retrospective cohort and found that 3GC-R increased the hazard of death when compared to either 3CG-S organisms (HR 1.63, 95%CI 1.13–2.35) or when compared to non-infected patients (HR 2.88, 95%CI, 2.22-3.74)[108]. The second was a study of neonates in Taiwan, though it is unclear if the patients were recruited prospectively or retrospectively. This study found a higher mortality when compared to non-infected controls (OR not reported) but not when compared to 3GC-S BSI OR 1.48; 95% CI .64 to 3.45)[109].

The majority of other studies use 3GC-S BSI alone as the comparator group, again finding significant association of resistance on mortality in multivariable models (Appendix Table 1.1). Four studies found no effect of 3GC-R on mortality, but all of these were retrospective cohorts.[27, 110-112].

1.4.3 Health economic burden from 3GCR-E BSI

1.4.3.1 Overview

The funding and resources allocated to mitigate the impact of AMR are huge[1, 113], and need to be backed by accurate enumeration of costs in order to understand the efficiency of these investments and incur the appropriate policy responses in terms of setting priorities for treatment and prevention of DRIs. These economic considerations are of particular importance where resources are limited[114] and the O'Neill Report predicted that the economic cost of AMR to healthcare will be highest in sSA[1].

Healthcare economic burden studies commonly investigate the cost of illness to either healthcare providers, measured at the level of an individual institution, or healthcare users, measured at patient level[115-117]. Direct measurement of costs to the healthcare institution include hospital charges, staff salaries, and resources such as drugs and investigations. Costs to patients include direct expenditure incurred during a hospital stay such as transport to hospital, and the indirect costs such as loss of income from time off work. In AMR terms, these costing approaches can be applied to the cost of drug-resistant infections compared to drug susceptible infections

(i.e. the additional cost of treating an infection in the presence of resistance)[116]. Intuitively, DRIs incur additional costs resulting from the requirement for more expensive antibiotics, investigations and procedures and longer hospital stays[118].

Hospital length of stay (LOS) is considered a key driver of hospital costs[119] and is therefore included as a variable in economic models used to calculate cost of hospital admission[120, 121]. A review of the impact of 3GCR BSI on hospital LOS is therefore included in this section, but in addition to being useful in economic calculations, LOS is frequently used as a proxy outcome measure of patient morbidity. Other morbidity outcome measures such as ICU admissions, need for surgery, re-hospitalisation and disability status at discharge or follow-up, are more variably used, and not always applicable to low-income settings[116].

1.4.3.2 Search strategy

The peer-reviewed literature was searched for studies which investigate the cost of AMR in terms of length of stay, patient costs or direct hospital costs. Search terms included: (MeSH search terms (Drug resistance OR Antimicrobial resistance OR Bacterial resistance OR Drug Resistance, Bacterial) AND text word search (antibiotic resistan* OR antibacterial resistan* OR antimicrobial resistan* OR antimicrobial drug resistan* OR antibiotic drug resistan* OR antibacterial drug resistan* OR bacteraemia OR bacteremia OR bloodstream inf*) AND MeSH search terms(Health Care Costs OR Cost of Illness OR Hospital Costs OR Length of stay)) OR text word search(economic).

A search pertaining to sSA produced only two relevant articles and was therefore expanded with no country restrictions, in order to get a full understanding of appropriate study methodologies. Included studies had to clearly delineate BSI from other infections and 3GC-R E from other bacteria/antibiotic combinations.

1.4.3.3 Hospital length of stay

Appendix Table 1.2 shows data from 13 studies investigating attributable LOS from 3GCR-E BSI. Only two studies were from sSA, both from Senegal[92, 103].

LOS estimates varied, from no attributable increase (one study), to 11 days excess hospital stay attributable to 3GC-R. Studies are not comparable due to a mixture of retrospective and prospective designs, variety of settings (ITU, acute admissions or hospital-wide) and the different analytic models used (linear regression, multistate models, cox regression). The first study from Senegal used a retrospective cohort design, comparing patients of all ages with 3GC-R and 3GC-S BSI. Using a multistate model, they estimated that 3GC-R was associated with an excess LOS of 4.3 days (95% CI, 3.8 to 4.6)[103]. The second study from sSA, also from Senegal, used a case-control design in a neonatal unit, estimating an excess LOS of 7.9 days (95%CI, 7.6-9.2) attributable to 3GC-R[92].

Studies aiming to understand attributable LOS from AMR are at risk of time-dependent bias, if the entire length of hospital stay is compared between groups of patients with susceptible and resistant infection (i.e. if

all infections are assumed to have occurred at the time of admission)[122]. Studies that do not control for this, will overestimate attributable LOS and costs by including all bed-days and costs that are incurred both before and after the infection occurred. [119, 123, 124].

The most effective way of accounting for time-dependent bias is use of multistate models, which include time from admission to occurrence of BSI as a time-dependent intermediate exposure variable and discharge or death as the final absorbing state[108, 125, 126]. This approach, which can be used in both regression and survival models, tends to produce more conservative LOS and economic impact estimates by avoiding bias and overestimation of time-dependent exposures[124]. Other methods, which are considered less optimal in terms of reducing bias[124], include adjustment at the design phase (matching groups of patients based on hospital LOS prior to BSI onset) or in the analysis phase, by including time to BSI as a regression covariate[121]. Only four studies, including the two from sSA, used a multistate model to control for time-dependent bias[92, 103, 108, 125, 127], the remainder using a matching approach[118, 128] or not controlling for this bias at all (Appendix table 1.2).

1.4.3.4 Direct hospital costs

Direct hospital costs of a DRI generally comprise two major components: the consumables used to treat the infection (e.g. antibiotics, investigations and procedures) and cost of days of admission[120]. Primary resource-based costing studies (to estimate unit costs for each hospital resource) are first required before total direct health provider costs of consumables can be estimated[129]. Cost of days of admission requires an estimate of cost per

day of hospital care, accounting for staff salaries and cost of running central support services such as pharmacy/radiology[129, 130]. Cost per day of hospital care (the economic value of a bed-day) can then multiplied by quantity of additional bed-days required for DRI, for inclusion in total hospital costing models[119, 129].

Nine studies investigated the economic impact of 3GC-R from the hospital perspective (Appendix Table 1.2) reporting either total hospital costs, or hospital charges. Total additional costs of 3GC-R in studies of single-sites ranged from no additional cost in a German cohort[128], to an additional $29,379 per admission in a cohort from the USA[118]. Most studies report total additional costs over a particular time period, rather than costs per hospital admission, though in many cases it is unclear what is being presented and cost estimates are provided variably at the level of individual or multiple institutions. Only four studies, including one from sSA reported a clear breakdown of which costs were included in the analysis (e.g. drugs, investigations, staff costs)[103, 108, 128, 131]. The single study reporting from sSA, estimated an additional total hospital costs of + EUR100 (95%CI, EUR78 to EUR125), with hospital bed-days and antibiotics contributing to the majority of these costs[103]. None of the identified studies reported their primary resource-based costing methods.

As with LOS estimates, the heterogeneity in study design, setting and analytical approach prohibits a pooled estimate of the economic burden of 3GC-R BSI. The majority of studies do, however, report a significant impact of 3GC-R on direct hospital costs (Appendix Table 1.2).

The absence of data from LMICs is striking, given the high burden of disease[16] and likely significant impact on healthcare systems.

1.4.3.5 Patient costs

A second approach taken by cost of illness studies, is to measure patient level costs, which include the direct non-medical costs associated with an admission, such as cost of transport to hospital, toiletries and food and the indirect costs or productivity losses, namely loss of income caused by the need for hospital admission, premature death or disability.

Literature search identified only one study which considered patient level costs in the context of AMR. This study used collected mortality data from Thailand and the USA and converted these into productivity losses by multiplying total life years lost by Gross Domestic Product (GDP) per capita. For 3GC-R *E. coli*, annual productivity losses were estimated at 32.1 million US$ (Thailand)and 360.8 million US$ (USA)[113]. No studies of patient level economic costing from sSA were identified but this approach has been taken in Malawi in the context HIV admissions[115].

1.4.3.6 Societal costs

Concentrating on the economic impact on a healthcare institution or an individual may underestimate the full economic burden on society[132]. Total societal costs per individual can be modelled using the costing methods described above, by summing total direct heath provider costs, direct medical and indirect costs[115]. On a national scale, societal estimates of patient and hospital costs associated with AMR in have been

estimated at €1.5 billion in 2007 in Europe[133] and $55 billion in 2000 in the USA[134]. These estimates incorporate both direct hospital direct costs and at patient level, lost earnings resulting from illness or premature death. However, the morbidity and mortality burden data used to calculate these estimates were very limited, coming from small and mostly retrospective studies and clearly, this approach in LMICs would suffer from lack of data to underpin the estimates.

Another approach, the one taken in the O'Neill report, is to make projections on the "full potential economic costs" of AMR i.e. the societal costs of a scenario where no antibiotics work, accounting for more than those related to the healthcare sector alone. These macroeconomic models account for all losses of modern health care that would occur from untreatable infections, including wider indicators such as national income and economic growth[116, 135]. Such models have resulted in estimate of between cumulative projected cost to the global economy of 100 trillion USD by 2050[1]. This report estimates that LMICs will suffer the worst economic losses, but again, there is lack of data to back these projections[71].

1.5 Patient quality of life outcomes as a measure of healthcare burden

Patient related outcomes, including health related quality of life (HR-QoL), are increasingly recognised as an important measure of health burden[130]. HR-QoL measures are frequently incorporated into cost evaluations that aim to expand their assessment of health burden, by going

beyond traditional clinical metrics such as mortality, to measure health gain in terms of both quantity and quality of life[136].

The most widely used generic measures of health gain, include disability-adjusted life years (DALYs) and quality-adjusted life years (QALYS) which are useful summary measure of health burden that can be used to make comparisons between different diseases and health states[137]. The WHO have proposed AMR-specific indicators for monitoring Sustainable Development Goals (SDGs) and suggest DALYs or QALYs as a key metric that will allow for comprehensive translation of AMR into patient burden[138].

QALYs incorporate the impact of disease on a patients length and quality of life (i.e. are a measure of healthy years lived), and are calculated by multiplying duration of time spend in a health state (e.g. survival) by a HR-QoL weight such as the Euro-Qol EQ-5D[130]. DALYs, do not incorporate HR-QoL per se, but measure loss of health by summing years of potential life lost and years of productive life due to disability (1 DALY = 1 year of healthy life lost)[136].

DALYs are widely used in GBD estimates[139], but literature review identified only one study that applied one of these composite health measures to antibiotic resistant bacteria[140]. This large consortium used European surveillance data to estimate the burden of a number of key antibiotic-bacterium combinations and expressed this burden in DALYs. The study estimates 170 DALYs per 10000 population from all included antibiotic resistant bacteria, with over one-third of these attributable to

3GC-R *E. coli* and *Klebsiella*. Estimates showed marked heterogeneity between countries, but the overall burden was similar to the burden of influenza, TB and HIV combined[140]. Although not yet applied to AMR, this general HR-QoL approach has been taken in Malawi using QALYs as health outcome measure to assess cost-effectiveness of HIV self-testing[115] and to assess QoL in HIV-infected hospitalised patients[129].

1.6 Assessing the burden of AMR and rationale for thesis

Although data are limited, the prevalence of gut mucosal carriage of 3GCR-E in community populations in sSA is high. Studies frequently link 3GCR-carriage with hospitalisation or other healthcare access, but there are currently too few data on community and household level risk factors to make firm conclusions about the drivers of acquisition and transmission of these organisms. For this, detailed longitudinal studies from healthy community populations are needed.

Rates of invasive disease with 3GCR-E in sSA, approach some of the highest in the world, but the impact of these infections on patients and health systems is currently unknown. The largest African AMR surveillance dataset comes from Malawi, where 20 years of blood culture surveillance has shown a rapid rise in prevalence of 3GC-R amongst Enterobacterales. Despite these alarming trends, the impact of these infections on patients and health-systems is currently unknown.

Microbiological laboratory data alone are clearly an insufficient measure of AMR burden and comprehensive assessments will need to link prevalence data to patient and health-system outcomes. Health burden indicators, such as mortality are important and a useful way to convey the impacts of AMR to policy makers and clinicians, as well as to make comparisons between the health impacts of DRIs and other diseases[141].

The O'Neill Review used data from the European Centre for Disease Control (ECDC) and Centre for Disease Control and Prevention (CDC) surveillance reports, to generate pathogen specific mortality rates from AMR. However, this report assumed that mortality from DRIs, in countries where data were lacking, was equal to that observed in the USA[1, 71, 141]. The major limitation of this review was therefore the lack of data from LMICs[1] but there were additional methodological challenges with the studies incorporated into the final estimates of this report. The data used to calculate mortality did not account for competing outcomes (i.e. that some patients with DRIs are discharged alive) or for time dependent bias (infections are sometimes acquired post admission) and therefore may have overestimated the mortality, morbidity and cost attributed to AMR[71]

The appropriate study designs and statistical inference methods used to quantify the mortality burden from AMR have recently been debated [141]. This consortium reflected on the challenges of attributing mortality to a DRI and the lack of a universally accepted method was acknowledged. One approach, taken by many of the studies I have described, is to estimate "attributable mortality", using the counterfactual assumption that death

would not have occurred if the organism causing the infection had been drug-susceptible (or if there had been no infection). This assumption allows for comparison between patients with resistant and sensitive bacterial infection (or between patents with resistant infection no infection). The Global Antimicrobial Resistance Surveillance System (GLASS), launched by WHO are developing a standardised approach to AMR burden data collection and analysis, recommend the use of this "attributable mortality" approach[142].

Regardless of comparator group, these models must adjust for appropriate confounders such as comorbidities and patient age[143]. There are disadvantages to the attributable mortality approach, not least that it is not easy to standardise, with studies using different comparator groups, populations and analytic models[16, 141]. However, other approaches, such as measuring all-cause mortality without a comparator group, are more likely to overestimate AMR burden[141].

A comprehensive approach to measuring the health burden of AMR should not report on fatal outcomes only, but should also incorporate an assessment of loss or gain of healthy life years or quality of life, using comparable summary measures such as DALYs or QALYs[140]. Economic cost evaluations are of particular importance in LMICs, which have multiple, competing priorities for limited healthcare resources and budgets.

The majority of studies describing AMR burden do not separate community from hospital acquired infections[16, 141]. Knowledge of these origins is essential, however, since interventions targeted at reducing

nosocomial DRIs will likely be very different to those targeted at reducing community acquired infection. This information can be difficult to collect in retrospectively acquired data, but prospective studies should aim to make these distinctions.

Finally, standardised and quality assured microbiological laboratory methods are needed to generate accurate AMR burden estimates. A recent publication from the AMR surveillance community aimed to identify exactly what data are minimally acceptable and what data are ideal, to produce useful AMR prevalence estimates[144]. Studies should include reporting of quality assurance, use internationally standardised methods and provide a clear account of the microbiological sampling criteria, study or surveillance sampling frame and the laboratory methods used to generate resistance data[16, 144].

Although the gold-standard methodologies to estimate AMR burden are still debated[141], it is clear that studies should link quality assured laboratory results to the clinical and economic outcomes, using prospectively collected data and should separate hospital and community acquired burden. It is also clear that these data from sSA, are severely lacking but urgently needed.

1.7 Study approach

The study described in this thesis was designed to investigate the attributable morbidity, mortality and economic cost of third-generation cephalosporin resistant bloodstream infections in Malawi, a country which

has the largest bacteraemia and AMR surveillance dataset from sSA, but in which the health burden of AMR infections is currently unknown. I aim to address this knowledge gap by assessing the healthcare burden of resistance to one of the most commonly used and frequently last-line antibiotics in hospitalised inpatients in Malawi.

Using prospective recruitment and detailed characterisation of all BSI episodes, I will describe the impact of bloodstream infection on patients, and in turn, the burden of 3GCR-E infection. In line with recently published guidance on quality reporting of AMR data [142, 144], and taking advantage of the quality-assured Malawi-Liverpool Wellcome Trust (MLW) laboratory, I am be able to provide a clear account of microbiological sampling criteria, sampling frame and laboratory methods, as well as clinical metadata including empiric antibiotic regimens, HIV status and healthcare attendance.

In addition to measuring attributable mortality from 3GCR-E BSI I will use hospital LOS as a measure of patient morbidity. Further, by measuring the incremental cost of resistant compared to sensitive infections, I will use costing tools which have been previously adapted for use at QECH[115], to measure direct hospital costs, direct non-medical and indirect patient level costs and finally health related QoL estimates for patients with 3GC-R and 3GC-S BSI.

The second part of this study is designed to accurately assess household level risk factors for gut mucosal colonisation with 3GCR-E, based around a randomly selected longitudinal cohort of households in urban Blantyre.

1.7.1 Hypotheses

1. Bloodstream infection with 3GCR-E carries a high attributable morbidity and mortality in Malawi and places a significant economic burden on healthcare provision and on patients, when compared to BSI with 3GCS-E.

2. Community acquisition is driving the high prevalence of 3GC-R *E. coli* carriage in humans in urban Blantyre, and the diversity in human 3GC-R *E. coli* carriage can be explained by household level determinants.

1.7.2 Specific Aims

The specific aims of this thesis are:

1. To investigate the association between 3GCR-E BSI and death and 3GCR-E BSI and hospital length of stay in hospitalised patients at Queen Elizabeth Central Hospital, Blantyre Malawi;

2. To compare and contrast the healthcare provider costs, patient costs and healthcare related quality of life outcomes amongst adult patients hospitalised with 3GCR-E BSI and 3GCS-E BSI;

3. To describe the associations of 3GC-R *E. coli* gut mucosal colonisation in urban Blantyre and using longitudinal microbiological surveillance of humans and investigation of household level risk factors.

1.8 Thesis overview

This thesis describes two longitudinal cohort studies based in Malawi, the methods for which are described in Chapter 2. The first is a hospital cohort of patients who have BSI with Enterobacterales or *Acinetobacter* spp.. Chapter 3 presents the clinical characteristics and microbiology of the cohort, alongside exploratory modelling of risk factors for 3GC-R BSI. Chapter 4 presents the outcomes of participants in the cohort and analyses the impact of 3GC-R on in-hospital mortality and hospital length of stay, while Chapter 5 analyses the impact on economic costs. The second cohort is a household study which cultures 3GC-R *E. coli* from healthy individuals in urban Blantyre. Chapter 6 presents the results of longitudinal models designed to understand risk factors for 3GC-R gut mucosal colonisation in these individuals and Chapter 7 uses these data to develop dynamical transmission models of 3GC-R carriage. Finally, Chapter 8 provides conclusions and suggests future research questions.

1.9 Appendix

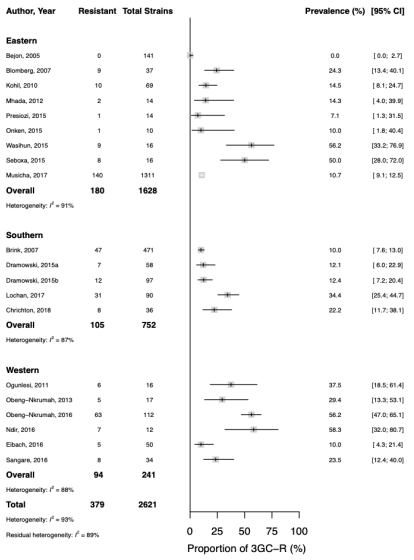
1.9.1 Search terms for 3GC-R gut mucosal carriage review

(ESBL OR Extended-spectrum beta-lactamase OR third-generation cephalosporin) AND ((Angola OR Benin OR Botswana OR Burkina Faso OR Burundi OR Cameroon OR Cape Verde OR Central African Republic OR Chad OR Comoros OR Republic of the Congo OR Congo Brazzaville OR Democratic republic of the Congo OR Cote d'Ivoire OR Djibouti OR Equatorial Guinea OR Eritrea OR Ethiopia OR Gabon OR The Gambia OR Ghana OR Guinea OR Guinea-Bissau OR Kenya OR Lesotho OR Liberia OR Madagascar OR Malawi OR Mali OR Mauritania OR Mauritius OR Mozambique OR Namibia OR Niger OR Nigeria OR Reunion OR Rwanda OR Sao Tome and Principe OR Senegal OR Seychelles OR Sierra Leone OR Somalia OR South Africa OR Sudan OR Swaziland OR Eswatini OR Tanzania OR Togo OR Uganda OR Western Sahara OR Zambia OR Zimbabwe)OR Africa)).

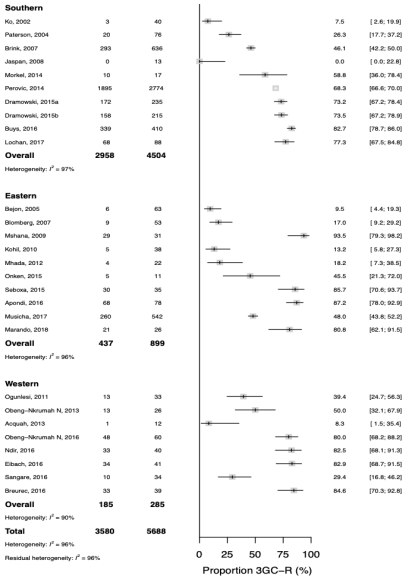
1.9.2 Search terms for 3GC-R prevalence and mortality review

MeSH search terms (Drug resistance OR Antimicrobial resistance OR Bacterial resistance OR Drug Resistance, Bacterial, OR Microbial Sensitivity Tests OR Drug Resistance, Microbial) AND text word search (antibiotic resistan* OR antibacterial resistan* OR antimicrobial resistan* OR antimicrobial drug resistan* OR antibiotic drug resistan* OR antibacterial drug resistan* OR bacteraemia OR bacteremia OR bloodstream inf*) AND (Angola OR Benin OR Botswana OR Burkina Faso OR Burundi OR Cameroon OR Cape Verde OR Central African Republic OR Chad OR Ivory Coast OR Cote d'Ivoire OR Congo OR Comoros OR Djibouti OR Equatorial Guinea OR Eritrea OR Ethiopia OR Gabon OR Gambia OR Ghana OR Guinea OR Guinea Bissau OR Kenya OR Leshoto OR Liberia OR Madagascar OR Malawi OR Mali OR Mauritania OR Mozambique OR Namibia OR Niger OR Nigeria OR Rhodesia OR Rwanda OR Sao Tome and Principe OR Sengal OR Seychelles OR Sierra Leone OR Somalia OR South Africa OR Sudan OR Swaziland OR Tanzania OR Togo OR Uganda OR Zambia OR Zimbabwe OR Africa))

A

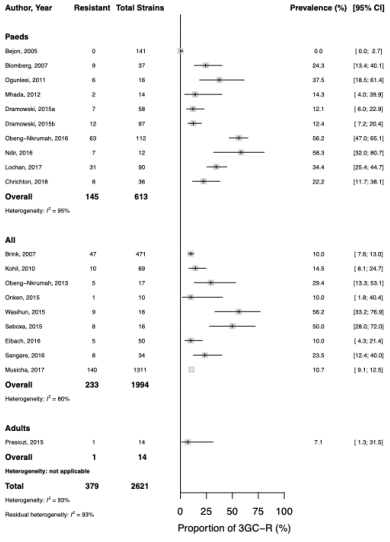


B

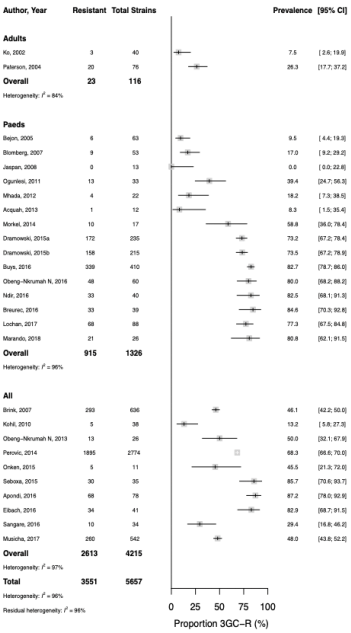


Appendix figure 1.1 (A) Forest plot of 3GC-R *E. coli* with subgroup analysis by Africa region. Test for subgroup differences (random effects model) Q=3.09, d.f = 2, p=0.21 and (B) Forest plot of 3GC-R *Klebsiella* spp. with subgroup analysis by Africa region. Test for subgroup differences (random effects model) Q=0.46, d.f = 2, p=0.79. Modified from[16].

A



B



Appendix figure 1.2 (A) Forest plot of 3GC-R *E. coli* with subgroup analysis by age group of population and (B) Forest plot of 3GC-R *Klebsiella* spp. with subgroup analysis by age group. Test for subgroup differences not done due to small number of studies in adult subgroups. Modified from[16]

Appendix Table 1.2 Studies reporting attributable length of stay or excess hospital costs from 3GCR-E BSI

First author, Year	Country	Study type, Setting	Comparator group(s)	LOS (timing)* Model	Cost drivers	Excess cost	Author conclusions
Blot, 2002	Belgium	Retrospective cohort, ICU	3GC-S BSI	LOS(admission and onset of BSI) NR	Not done	+8 days, P=0.33 (timed from onset BSI) +12 days, P<0.01 (timed from admission)	No association between ESBL and longer stays if timed from onset of BSI
Cosgrove, 2002	USA	Nested matched cohort, Hospital-wide	3GC-S Enterobacter	LOS (BSI onset) MVLR	Per patient hospital charges (no breakdown)	+ 9 days, p<0.001 + $29,379, P<0.001	3GC-R in Enterobacter is associated with adverse cost outcomes
Lambert, 2001	European multisite	Prospective cohort, ICUs	3GC-S BSI *E. coli*	LOS in ICU Time-dependent regression	Not done	HR 1.09 (95%CI, 0,43-0.89) Absolute no. of days NR	No additional effect of 3GC-R

Appendix Table 1.2 Studies reporting attributable length of stay or excess cost from 3GCR-E BSI (continued)

First author, year	Country	Study type, Setting	Comparator group(s)	LOS(timing)* Model	Cost drivers	Excess cost	Author conclusions
Leistner, 2014	Germany	Case-control, Hospital-wide	3GC-S BSI *E. coli*	LOS (BSI onset) MVLR	Hospital costs: Bed days Staff time Drugs Procedures	+ 6 days, P=0.39 (adjusted) + $1,479, P=0.359 (adjusted)	No significant economic differences between 3GC-Rpositive and negative infections
Melzer, 2007	UK	Prospective cohort, Hospital wide	3GC-S BSI *E. coli*	LOS (BSI onset) NR	Not done	+ 3 days, P=0.11	No difference in LOS in patients who survived, measured from time of onset of BSI.
Ndir, 2016	Senegal	Retrospective cohort	3GC-S *E. coli*	LOS (NR) MSM	Direct hospital costs: drugs investigations bed days	+ 7.9 days (95% CI, 7.6-9.2) EURO 100 (95% CI 78-125)	Significant economic impact of 3GC-R infection

Appendix Table 1.2 Studies reporting attributable length of stay or excess cost from 3GCR-E BSI (continued)

First author, year	Country	Study type, Setting	Comparator group(s)	LOS(timing)* Model	Cost drivers	Excess cost	Author conclusions
Ndir, 2016	Senegal	Case-control	3GC-S BSI	LOS (NR) MSM	Not done	+ 4.3 days	3GC-R associated with prolonged LOS
Schwaber, 2006	Israel	Retrospective cohort, All wards	3GC-S BSI	LOS (NR) MVLR NR	Per patient hospital charges (no breakdown)	+ 6 days, P<0.001 (unadjusted) OR 1.56 (adjusted) + $9,620 (adjusted)	3GC-R is a significant predictor of increased LOS and increased cost
Stewardson, 2016	Europe, multicentre	Retrospective cohort, Acute admissions	3GC-S BSI	LOS (NR) MSM	Hospital costs: bed days hospital services (no breakdown)	+ 4.89 days (2.08-4.98) + EUR 250 (60-1100)	3GCR increases costs and LOS
Tumbarello, 2010	Italy	Retrospective cohort, Adult wards	3GC-S BSI	LOS (NRO NR	Hospital costs (no breakdown)	+ 7 days, P=0.02 +EUR 5,026, P=0.04	3GC-R BSIs are associated with longer and more costly hospital stays

Note:

* Timing refers to onset of measurement for LOS calculation

MSM = multistate model, MVLR = multivariable logistic regression, NR = not reported,

Chapter 2

General methods

2.1 Chapter overview

This chapter describes the general methods for the study from which this thesis is derived. Here, I describe the clinical, epidemiological and laboratory methods which are common to all analyses. The subsequent results chapters contain the specific methods relevant to the particular analysis they describe.

2.2 Study overview

The ANTIDOTE Study (Antimicrobial resistance study to determine outcomes and transmission of ESBLs) was a prospective observational study which recruited two cohorts. Cohort 1 (hospital) was designed to measure the attributable mortality, morbidity and cost burden from 3GCR-E BSI. Cohort 2 (household) was designed to measure determinants of acquisition and transmission of 3GCR-E gut mucosal carriage. Cohort 1 recruited participants from QECH, Blantyre, between January 2018 and March 2020. Cohort 2 recruited participants living in households in urban Blantyre, between April 2018 and November 2019.

The protocol for the hospital study (Cohort 1) is published and the questionnaires are available as extended data to the publication[145].

2.3 Study sites

2.3.1 Malawi

Malawi is a small, land-locked country in South-Eastern Africa (Figure 2.1), with an estimated population of 17.5 million people[146]. Classified as low-income by the World Bank, Malawi ranks 149th out of 205 economies (2018 GDP of USD 7.065 Billion)[147]. The Human Development Index (HDI), a composite index of life expectancy, schooling and per capita income, ranked Malawi 172nd out of 189 countries and territories in 2018, with 70.3% of the population estimated to be living on less than 1.9 USD per day[148].

Life expectancy at birth is estimated at 64.3 years[146]. Malawi was one of 10 low-income countries to reduce its under-five mortality by at least two-thirds between 1990-2018, but infant and neonatal mortality remain high, at 38 and 22 per 1000 live births. Malawi is meso-endemic for malaria, with those living on the shores of Lake Malawi or in the Shire River Valley considered at highest risk[149]. The HIV epidemic is generalised, with a prevalence in adults of 9.2% in 2018 and an estimated 74,000 children aged 0-14 living with HIV[148]. WHO 90-90-90 targets envisaged that by 2020, 90% of people living with HIV would know their status, 90% of people who know their HIV-positive status would be accessing ART and 90% of people on treatment would have supressed viral load[147]. In 2019, in Malawi, 90% of people living with HIV knew their status,79% of all people living with HIV were accessing ART and 72% of all people living with HIV were virally suppressed[148].

Malawi has a sub-tropical climate, with one annual rainy season, usually from November through to April. Blantyre city, is the second city and commercial capital of Malawi, with a population of 800,264. It is located within Blantyre district, population 995000 (2018 census), at an altitude of 1000m.

Figure 2.1: Map of Malawi, highlighted in yellow, showing Blantyre in the Southern region. Produced using openstreetmap.org, used under Creative Commons Attribution ShareAlike 2.0 licence CC-BY-SA.

2.3.2 QECH

Cohort 1 recruited participants from QECH, the largest government hospital in Malawi. QECH provides free healthcare to Blantyre and the surrounding districts, plus tertiary care to Malawi's Southern region. It has 1300 beds, frequently operating above capacity. Adult patients attending QECH must be referred from a primary health centre, or from another hospital if tertiary care is being sought. Patients are triaged 24-hours a day in the Adult Emergency and Trauma Centre (AETC) and either

discharged or referred to the appropriate speciality team (medicine, surgery or obstetrics and gynaecology). The medical and paediatric wards were the main source of patients recruited to this study, but patients were occasionally recruited from the surgical and obstetrics and gynaecology departments and from the Intensive Therapy Unit (ICU). This mixed-specialty 4-bedded ITU has facilities to operate a maximum of three ventilated beds, but problems with equipment maintenance and supply frequently limit this capacity.

The Department of Medicine runs two single-sex medical wards, each with an official capacity of about 60 beds, but these wards are frequently oversubscribed with patients nursed on floor mattresses between bed spaces and in corridors. The department also includes one mixed-sex TB ward and two six-bedded high-dependency units (HDUs). Medical care on these wards is basic and antibiotic supply inconsistent, with frequent stock-outs. At the time of writing, oxygen concentrators were available on the HDUs and oxygen cylinders are sometimes available but supplies are again inconsistent. Consultant physician reviews generally occur twice-weekly on the general wards and twice-daily on the HDUs. Medical input at other times is more variable and is conducted by junior doctors or medical students. Weekend and night time input for adult medical inpatients is minimal.

The Paediatric Department consists of a separate Accident and Emergency (A&E), neonatal unit (Chatinkha Nursery), general paediatric ward, oncology ward and a nutritional rehabilitation unit for children over 6 months of age who meet WHO defined Severe Acute Malnutrition criteria

(Moyo House). Children do not have to be referred from elsewhere and are triaged in A&E, other than outborne neonates who are usually seen directly on Chatinkha nursery. If admission is required, children are admitted to Paediatric Special Care (over 6 months of age), Paediatric Nursery (previously discharged neonates to infants up to 6 months of age) or to the relevant specialty ward. Paediatric ward rounds are conducted daily, with consultant rounds twice-weekly, and critically unwell children receive twice-daily consultant review. Medical input at night is limited. A slightly wider range of antibiotics is available on the paediatric wards, often made available through donations or private funding, but supplies are again variable. Oxygen is available on Paediatric special care and Continuous positive airway pressure (CPAP) on Chatinkha Nursery.

In July 2017 (prior to the start date of this study), the Mercy James Centre (MJC) for Paediatric Surgery and Intensive care was opened as a separate 50-bedded building, operating as part of QECH. Benefiting from private funding, MJC receives approximately 1600 admissions per year and houses the country's only Paediatric intensive care unit (PICU). Children are usually triaged to MJC from A&E or Chatinkha nursery for surgical input, but can also be referred directly from other hospitals and health centres. Higher level care is available on the 8-bedded PICU which has facilities to mechanically ventilate children of all ages. Antibiotic supply is more consistent at MJC, with carbapenems and amikacin generally available, though these still remain subject to stock-outs.

Nurse-patient ratios on all wards at QECH are generally poor. Each medical ward, for example, may be staffed by 1-3 trained nurses and a

variable number of nursing students. Basic nursing care is supplemented by a patient's family member, friend or neighbour, who is referred to as a guardian.

2.3.3 Malawi-Liverpool Wellcome Trust

The Malawi-Liverpool Wellcome Trust Clinical Research Programme (MLW) was founded in 1995 and is an affiliate of the University of Malawi College of Medicine, the Liverpool School of Tropical Medicine (LSTM) and the University of Liverpool. MLW is located in the grounds of QECH and conducts collaborative research within the Malawian community, with the aim of benefiting health.

2.3.3.1 MLW diagnostic microbiology service

A diagnostic microbiology service, provided through MLW, was established in 1998. This service operates 7 days/week, providing free aerobic blood cultures and CSF analysis to adult medical and paediatric patients at QECH. From March 2018, this service was extended to the Department of Obstetrics and Gynaecology, with a limited number of blood cultures offered per month. Other departments may occasionally order blood cultures at special request.

Clinical blood culture protocols at QECH state that in adults, 7-10mls of blood should be taken in patients presenting to the emergency department with a fever (axillary temperature $> 37.5^{o}$C) or clinical suspicion of sepsis, severe sepsis or septic shock. In children, 1-2 mls of blood is taken in patients with non-focal febrile illness and a negative malaria test or in

children with malaria whose fever persists despite treatment. A blood culture is also recommended in all premature or febrile neonates who are admitted to the neonatal unit. In a busy hospital with constrained resources and limited alternative diagnostics, blood cultures are often done on patients who do not fulfil these criteria, at the discretion of the attending clinician. These patients will not be excluded from my analysis.

The absolute number of blood cultures collected fluctuates on an annual basis, but has approached 15000 per year since 2013. The most commonly isolated pathogens are non-typhoidal *Salmonellae*, S*almonella* Typhi and *Streptococcus pneumoniae* (estimated minimum incidence ≥300/year) followed by the other Enterobacterales, in particular *E. coli* and *Klebsiella* spp. (50-299/year)[3].

2.4 Eligibility criteria

Inclusion and exclusion criteria for both cohorts are shown in Table 2.1. Cohort 1 was initially intended to recruit only participants whose blood culture was positive for non-Salmonella Enterobacterales. However, during the set-up period for the study, it became clear that 3GC-R *Acinetobacter* spp. were an emerging problem at QECH, particularly amongst neonates. *Acinetobacter* are closely related to Enterobacterales, often sharing similar AMR profiles, and given their importance as a nosocomial pathogen, I decided to include these patients in the cohort.

Table 2.1: Eligibility criteria for the ANTIDOTE study

Cohort	Inclusion criteria	Exclusion criteria
Cohort 1 (Hospital)	Patient of any age whose blood culture is positive for non-Salmonella Enterobacterales or *Acinetobacter* spp.	Blood culture is positive for Salmonella enterica (any serovariant)
	Patient is an inpatient at QECH or can be contacted for admission or assessment	Patient is unable to provide informed consent and there is no representative to provide informed consent
		Patient speaks neither English or Chichewa
Cohort 2 (Household)	Households are located within urban Blantyre	No household members speak English or Chichewa
		Household is located outside urban Blantyre
		Household members are unable to provide informed consent and there is no representative to provide informed consent

2.5 Study visits and procedures for Cohort 1 (hospital cohort)

2.5.1 Recruitment and enrolment

For practical reasons, participants were recruited between Monday and Friday, 8am – 5pm. On these days, twice-daily reviews of the blood culture bench in the in the MLW microbiology laboratory were conducted by the study laboratory technician, to identify consecutive blood cultures which were positive for non-Salmonella Enterobacterales or *Acinetobacter* spp.. Blood cultures which became positive over a weekend were identified on a Monday morning. At this bench review, blood culture results were first

screened for any Gram negative isolates which were hydrogen sulfide (H2S) negative on the Triple Sugar Iron (TSI) test. This microbiological test is routinely carried out in the MLW laboratory and differentiates Enterobacterales based on their ability to ferment sucrose, lactose and glucose and to produce hydrogen sulphide (H2S)[150]. In general, this test separates Salmonella spp. (H2S positive) from non-Salmonella Enterobacterales (H2S negative) and is done using a cultured isolate colony, 24 hours before a final organism identification is available. At this stage, the clinical team would screen patients whose blood culture was H2S negative. Full recruitment did not take place until the final culture result was known, usually 24-hours later. *Citrobacter freundii* and *Proteus* spp. can occasionally be H2S positive, therefore all final Gram negative identifications were checked on a daily basis.

Patients were approached for recruitment as soon as possible after the final blood culture result was known. If a patient had already been discharged by the time this result was available, they were contacted for review and potential recruitment if contact details were available. If a patient had died by the time the final result was available, the patient was recruited to the study and data-capture tools completed using a retrospective note review.

Figure 2.2 shows an overview of the study visit schedule and procedures for Cohort 1. At enrolment, following informed written consent, a baseline questionnaire was conducted, collecting demographic and clinical information including admission physiology, data on any pre-hospital healthcare attendance and a health-care utilisation survey. Information on antibiotics or other therapies administered up to that point were recorded.

Questionnaires were completed by a combination of patient interview, medical note review, and guardian interview if the participant was a child under 18 years of age, or was obtunded. If a patient had died by the time the blood culture was identified, the same baseline questionnaires were completed via medical note review, with data recorded as missing if was not available.

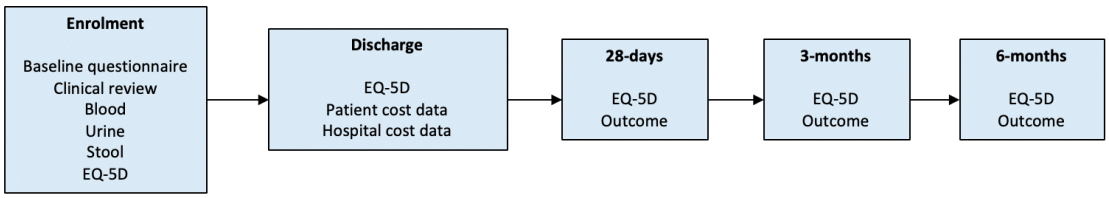


Figure 2.2: Overview of study visits and procedures for cohort 1 (hospital cohort). EQ-5D = EuroQoL quality of life questionnaire, explained in 2.5.3 below.

Measurement of vital signs was performed by the study nurse at enrolment. In adults, non-invasive blood pressure was measured using an automated cuff (Omron M2, Omron, Japan). Oxygen saturations were measured with a pulse oximeter (Contec CM50, Contec Medical Systems, China). Axillary temperature was measured using a digital thermometer (Omron FWH000, Omron, Japan). If adult patients were able to stand, they were weighed using mechanical floor scales (SECA261, UK) and measured using a mobile stadiometer (SECA213, UK).

In infants and children, tympanic temperature measured using a digital thermometer (MSRST613X, UK) and in children old enough, oxygen saturations were measured using the pulse oximeter. Neonates and infants were weighed using electronic baby scales (SECA365, UK) and measured

using an infant measuring mat (SECA 210, UK). Older children who were able to stand were weighed and measured using the adult equipment.

Each participant was assessed by a clinician (usually me, or a study clinician who then discussed each patient with me). At review, the admission history was noted and further clinical history and examination conducted as required. Treatment decisions were at all times at the discretion of the patient's clinical team. Antibiotics were administered following local guidelines (ceftriaxone as first-line for adults with sepsis, penicillin and gentamicin, or ceftriaxone for children). Where possible, the study team would help the clinical team to source appropriate antibiotics to treat the BSI, but antibiotics were not purchased or administered as part of the study.

2.5.2 Subsequent visits

Participants were followed up throughout their admission and the date of discharge or death was captured. Daily reviews were conducted by a study nurse on Monday to Friday, to capture details of any antibiotics administered and vital signs were repeated at each review. To allow for survival analysis, patients or their families were telephoned at 28-days, 90-days and 180-days, post enrolment. Follow up was conducted via telephone or with a visit to the participant's household if they could not be contacted by telephone. At each follow-up, details of any antibiotics received and healthcare visits since the previous contact were collected. If a participant had died by the time of follow-up, family members were asked for the date of death.

2.5.3 Health economic components

The methods for this component of the study are described in more detail in Chapter 5 and are summarised briefly here and in Figure 2.2. Primary costing studies to determine direct health provider costs of an admission to QECH have only been undertaken for adult patients, therefore children under 18 years of age were excluded from this component of the study.

Hospital costs: Upon discharge or death, information from the patient's medical record was extracted by a study clinician, to establish medications given, duration of hospital admission, types and numbers of investigations and procedures performed and the participant's outcome.

Patient costs (direct non-medical and indirect): A questionnaire was administered to patients and their guardians as as close to discharge as possible and ideally on the day of discharge if this could be anticipated. Data collected included cost of transportation, food, drinks, toiletries, clothing and other items bought during the hospital admission. For indirect costs, any time off work taken by participants or their guardians was recorded, together with self-reported income.

Health-related quality of life: The validated Chichewa version of the EuroQoL EQ-5D tool[115], was used to assess HRQoL of participants at recruitment and discharge from hospital and the descriptive components of the questionnaire were also used at Day 28, Day 90 and Day 180 follow-up[151].

2.5.4 Sample collection

Blood samples were collected from participants at enrolment and tested for full blood count (FBC), creatinine and CD4 if the participant was HIV-infected. Blood was collected aseptically using a vacutainer device as follows: serum for creatinine (adults, 4ml; children, 1.3ml; both tubes Grenier Bio-One, Austria) and EDTA for FBC/CD4 (adults, 4ml; children, 1.3ml; both tubes Grenier Bio-One, Austria). Point of care tests were carried out on capillary blood for capillary lactate (Lactate Pro 2, Arkray, Japan) and quantitative C-reactive protein (CRP) single test kit used with the NycoCard II Reader, Alere Technologies, Norway). HIV testing was organised as part of routine patient care, following Malawi national guidelines[152]. If a patient's HIV status was unknown at the time of recruitment, they were referred for HIV testing and counselling via standard QECH pathways.

Where possible, urine was collected from all participants at enrolment into a polypropylene universal container (Alpha Laboratories, UK), either directly by the participant or using a disposable bedpan. If the patient was catheterised, urine was collected directly from the catheter using a syringe. In children who were not toilet trained, a clean-catch urine was attempted by capturing a midstream urine into a universal container. All urine samples, except catheter specimens, were first screened using a dipstick (Siemens, UK). All catheter specimens and samples which were positive for leucocytes or nitrites on dipstick, were sent for culture (see Section 2.7, Laboratory methods).

One stool sample was taken from each participant at enrolment. Stool was collected into a sterile polypropylene universal container (Alpha Laboratories, UK). Rectal swabs were used if it was not possible for a patient to provide a bulk stool sample. These were taken by the study nurse using a sterile rayon-tipped swab (Technical Service Consultants Ltd, UK). Swabs were inserted into the rectum, rotated for 10 seconds and transported to the laboratory in Amies media.

All other sample collection, including urinary lipoarabinomannan (LAM), CSF and Sputum Xpert were done at the discretion of the clinical team providing routine care for the patient.

2.6 Study visits and procedures for Cohort 2 (household cohort)

2.6.1 Household selection and recruitment

Households were sampled using a spatially-weighted randomisation procedure which preferentially sampled areas of high population density. Within the Blantyre city polygon (Figure 2.3), the WorldPop 300m resolution population density raster was obtained. Then, for each of n households, a pixel was randomly selected either proportionally to the given population density with probability 0.9, or uniformly with probability 0.1.

The R code used to generate the sample points is shown in the appendix to this chapter and the sample points generated using this method are shown in Appendix Figure 2.1.

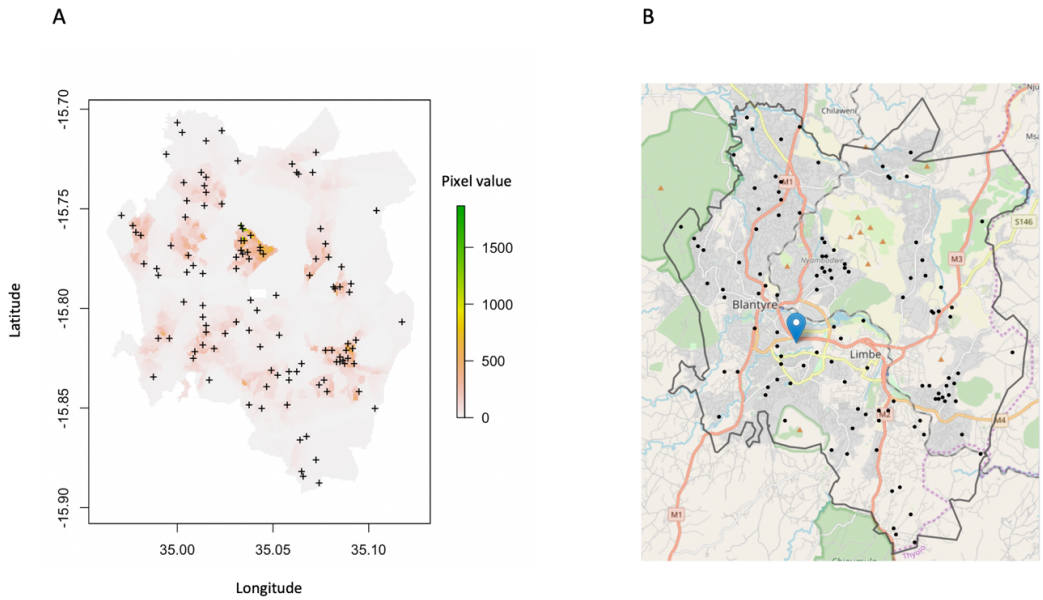


Figure 2.3: Household sample points generated using the mixture-distribution method. A shows the WorldPop raster for urban Blantyre, with sample points generated shown as black crosses. Pixel value is estimated number of people per pixel. Produced using WorldPop, used under Creative Commons Attribution 4.0 International License. B shows sample points superimposed onto map of Blantyre. Black polygon delineates outer boundary of urban Blantyre, sample points are shown as black circles and blue marker indicates location of QECH. Produced using openstreetmap.org.

To locate a household, field-workers were given a geo-spatial coordinate corresponding to a sample point, loaded onto a global positioning system (GPS) device (Garmin eTrex30, Garmin, UK), together with a pre-loaded route from MLW. Once as near as possible to the coordinate, field workers would spin a bottle and find the nearest household in the direction of the bottle. If the members of the nearest household members were unavailable or declined consent, a second house in the same direction was approached. Up to five households in the direction of the first bottle spin were approached for consent and if no household consented, the bottle was respun at the original GPS point and the process repeated. If there were no households in the vicinity of the sample point, field workers would drive to the nearest houses to that point.

Approximately 3-4 households were recruited each week. For logistical reasons, it was not possible to choose a sample point to visit at random on a given day. Instead, a traditional authority area (TA) within urban Blantyre was chosen at random at the start of the week and sample points within that area visited over the course of that week (e.g. Week 1, Ndirande; Week 2, Likhubula…..). Blantyre's 27 TAs, are shown in the appendix to this chapter.

2.6.2 Procedures at the baseline household visit

Following informed consent taken from a household representative, field workers used questionnaires to collect data on covariates potentially associated with 3GC-R *E. coli* carriage. Two levels of questionnaire were used: a household questionnaire which collected data on household level covariates such as socioeconomic indicators and WASH facilities and a separate individual level questionnaire for each consenting household member, which collected individual level covariates such as medical history, antibiotic use and travel. Questions were asked directly to household members wherever possible and information taken from government issued family-held records, "health passports", if they were available and the participants agreed.

Stool samples were collected from as many consenting household members as possible. Participants were provided with biodegradable bowls and sample containers to facilitate stool sample collection. Field workers returned the following day to collect samples.

2.6.3 Procedures at the follow-up household visits

Households were followed up at day 28 and at day 90 and 180. At each visit, household and individual level questionnaires were repeated with updated information as required, for example, asking about any healthcare attendance or antibiotic use since the previous visit. Stool samples were collected from as many consenting household members as possible at each visit.

2.7 Laboratory methods

Results of diagnostic blood tests were fed back to the relevant clinical team as soon as they were available and placed in the participant's medical notes. The MLW laboratory adheres to UK National External Quality Assessment Service (NEQAS) quality control.

2.7.1 Laboratory diagnostics

2.7.1.1 Haematology and biochemistry

FBC, CD4 and creatinine were processed in the MLW blood-science laboratory following standard protocols. FBCs are automated (Beckman Coulter HmX Haematology Analyser, Beckman Coulter, USA) as were CD4 cell counts (Becton Dickinson FACSCount, Becton Dickinson, USA). Serum was allowed to settle upright for 30-60 minutes and then centrifuged at 1300g for 10 minutes. Biochemistry testing for creatinine was then carried out (Beckman Coulter AU480 Chemistry Analyser, Beckman Coulter, USA).

2.7.1.2 Aerobic blood culture and antimicrobial susceptibility testing

Blood was inoculated into a single aerobic bottle and incubated using the automated BacT/ALERT system (bioMerieux, France). Incubated bottles which flag positive are processed with Gram stain and subculture. Enterobacterales and *Acinetobacter* spp. are identified to species level using Analytical profile index, API 20E (bioMerieux, France).

Before March 2019, antimicrobial sensitivity testing (AST) in the MLW laboratory, was carried out as per British Society of Antimicrobial Chemotherapy (BSAC) guidelines[153], and from March 2019, as per European Committee on Antimicrobial Susceptibility Testing, EUCAST guidelines[154].All blood culture isolates for participants recruited to the study were regrown by the study laboratory technician and AST carried out as per EUCAST guidelines [154, 155]as follows.

For each recruited participant the pure blood culture isolate was stored at -80°C on MicrobankTM bacterial storage beads (Pro-lab diagnostics, UK). Blood culture isolates were regrown from beads at 35°C for 18 hours. The direct colony suspension method was then used to make a suspension of pure colony in 1ml 0.9% sterile saline solution, to the density of 0.5 McFarland turbidity standard. After vortexing, a sterile cotton swab was dipped into the suspension and streaked evenly onto Muller-Hinton agar (MHA), aiming for confluent bacteria growth. Antimicrobial discs were applied and the resulting AST plates incubated at 35°C for 18 (+/-2 hours). Zones of inhibition for each antimicrobial were measured to the nearest millimeter and susceptibility categories interpreted according to

EUCAST breakpoint tables[155]. Table 2.2 shows the antibiotic discs used in the study.

For any Enterobacterales resistant to one or both of cefpodoxime or ceftriaxone on AST, ESBL production was confirmed using the species dependent combination disc method. The isolate was cultured overnight on MHA with discs of cefotaxime and ceftazidime (30 micrograms) with and without clavulanic acid (10 micrograms). For all organisms capable of carrying chromosomal AmpC, an AmpC-stable cephalosporin, cefipime (30 micrograms), was used with and without clavulanic acid (10 micrograms). ESBL production was confirmed if there was a difference of at least 5mm between discs with and without clavulanic acid. Plates were checked for confluent growth, by me, each time a new batch was made and in-house control strains used on each batch of agar.

Table 2.2: Antimicrobial discs used in AST for blood culture isolates

Enterobacterales	*Acinetobacter*
Co-trimoxazole 25 µg (SXT25)	Co-trimoxazole 25 µg (SXT25)
Gentamicin 10 µg (CN10)	Gentamicin 10 µg (CN10)
Ciprofloxacin 5 µg (CIP5)	Ciprofloxacin 5 µg (CIP5)
Meropenem 10 µg (MEM10)	Meropenem 10 µg (MEM10)
Amikacin 30 µg (AK30)	Amikacin 30 µg (AK30)
Chloramphenicol 30 µg (C30)	
Piperacilin-tazobactam (PTZ 36)	
Co-amoxiclav (AUG 30)	
Cefpodoxime (CPD10)	
Cefoxitin (FOX 30)	
Ceftriaxone (CRO30)	

2.7.1.3 Urine culture

Urine was cultured on commercially available selective chromogenic media, CHROMagar Orientation (CHROMagar, France), to obtain a presumptive organism identification. Urine was cultured overnight and if there was no growth after this time, a "no-growth" result was recorded. For cultures with growth, the number of bacteria were estimated using a colony counting method (one colony is approximately 1,000 cfu/mL (1x106 cfu/L). Below a recognised threshold (105 cfu/mL), the organisms grown were considered contaminants and the result reported as "no significant growth". Above the threshold it is more probable that a true urinary tract infection is occurring and presumptive identification using colour change on Orientation agar was carried out as per Table 2.3. Sweeps of pure growth on Orientation agar were stored on Microbank beads.

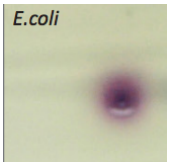
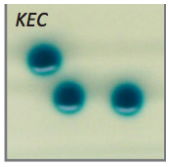
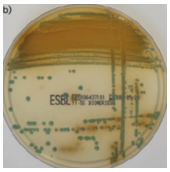
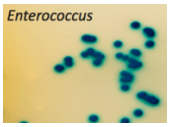
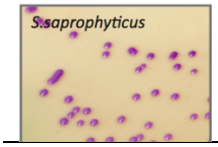
2.7.1.4 Selective stool culture

All stool and rectal swab samples were batch cultured on a weekly basis and stored at 4°C pending processing. Samples were plated onto ESBL selective chromogenic media, CHROMagar ESBL (CHROMagar, France) and cultured overnight at 37°C. Presumptive identification was then made based on colony colour changes which are similar to those for CHROMagar Orientation (pink = *E. coli*, blue = *Klebsiella* spp., *Citrobacter* spp., or *Enterobacter* spp., white = other species). Blue and white colonies and any colonies where the colour was doubtful, were speciated using API 20E. CHROMagar ESBL sufficiently sensitive for ESBL detection that confirmatory ESBL testing is not required[156]. An example of CHROMagar ESBL colony appearance is shown in Figure 2.3.

For Cohort 1 participants, 5 morphologically distinct colony picks grown from CHROMagar ESBL were stored separately in 5 separate Microbank vials. In addition, a sweep of growth from the ESBL selective plate was stored. From January 2019, stool was also cultured on MacConkey agar, which selects for Gram negative Enterobacterales, but not for resistance. From this time, sweeps of growth from MacConkey agar were also stored.

For Cohort 2 patients, 5 morphologically distinct colony picks grown from CHROMagar ESBL were stored separately in 5 separate Microbank vials. In addition, a sweep of growth from the selective plate was stored.

Table 2.3: Typical colony appearance of urinary pathogens on CHROMagar Orientation

Appearance	Organism
Dark pink to reddish 	*E. coli*
Metallic blue 	Klebsiella /Enterobacter/Serratia/Citrobacter group
Brown halo 	Proteus/Morganella/Providencia group
Small colonies, turquoise blue 	*Enterococcus* spp.
Small colonies, light blue 	*Streptococcus agalactiae* (Group B Strep)
Small colonies, dark pink 	*Staphylococcus saprophyticus*

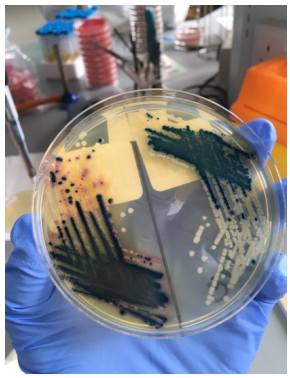


Figure 2.4: Typical colony appearance on CHROMagar ESBL. In this sample, pink colonies were *E. coli*, blue colonies were *Klebsiella* spp., and white colonies were *Acinetobacter* spp.

2.8 Statistical methods

All statistical analyses were conducted using R.Studio version 4.0.1 (R Foundation for Statistical Computing, Vienna, Austria)[157]. Details of individual analyses are included in the relevant chapters. R scripts for this thesis are available at: https://github.com/rlester1/thesis.

2.9 Sample size considerations

2.9.1 Cohort 1

The primary outcome for cohort 1 was to detect a difference in in-hospital mortality rates between participants with 3GC-R and 3GC-S bloodstream infection and the study was powered for this outcome.

A sample size calculation to inform the size of this hospital cohort was undertaken. There were insufficient existing studies from sSA powered to detect mortality from 3GC-R BSI, on which to guide our sample size estimates, but a large multi-centre European study found that mortality was 14% higher in patients who had an 3GC-R BSI versus those who had 3GC-S BSI[158]. Based on this, we aimed to recruit 250 patients to the cohort, which would provide 80% power to detect a difference in 28-day mortality rates of 10% vs. 24.1% at a 5% Type-I error rate. If this recruitment target was not achieved, a more modest 200 patients would provide 80% power to detect mortality of 10% vs 25.8%. These calculations assume a balanced design, i.e. a ratio of 1:1 for 3GC-R and 3GC-S infections, which was reasonable based on 2016 figures for Enterobacterales BSI from QECH[3].

As the study progressed, it was clear that a high proportion of patients had died by the time the final blood culture result was available. Although these patients were still recruited via retrospective note review, more comprehensive data and sample collection were available from patients who were alive at recruitment. I therefore decided to aim for recruitment of 250 participants who were alive at enrolment, in addition to the patients who had died.

2.9.2 Cohort 2

Cohort 2 aimed to recruit healthy humans in households, to determine risk factors for acquisition and carriage of 3GC-R *E. coli* using a transmission model approach.

Sample size estimation for parameter inference on dynamical transmission models is difficult, due to the large number of variables and potential unknowns that will be incorporated into the model. The sample size for Cohort 2 was therefore based on prevalence estimation of 3GCR *E. coli* carriage in the community.

At the time of designing the study, the only prevalence estimate available from Malawi was an unpublished cross-sectional survey suggesting that up to 50% of adult medical patients are colonised with 3GCR (K.Gray, unpublished). A sample of 300 stools would have 80% power to detect an 3GCR *E. coli* carriage prevalence of 50% (95% CI 35-65%). With an average of 4 people people per household in Blantyre, I therefore aimed to recruit a minimum of 75 households. The sample size was inflated to 110 to account for potential attrition of full households.

2.10 Data management

Data were collected and stored in accordance with International Conference of Harmonisation (ICH) Good Clinical Practice (GCP) guidelines and all members of the study team were GCP trained. Data were collected using a combination of Open Data Kit software (ODK) and Intelligent Character Recognition Scanning Data Capture Software (TeleForm, OpenText, Canada).

At the point of recruitment, each participant was given a unique identifier and data were anonymised. ODK forms were loaded onto Acer Iconia 32GB tablets and completed ODK forms pushed daily to the dedicated secure structured query language (SQL) database, managed by the MLW data team. Teleform paper forms were checked, scanned and validated by the MLW data team, in discussion with me and the clinical team if required. Validated TeleForm data were pushed to the SQL database in batches. All data on the study database were stored securely, with access restricted to the study PI and the lead data managers at MLW. Completed paper TeleForm records were stored securely in the MLW data department. Results of laboratory investigations were stored anonymously on the MLW PreLink laboratory information management system (LIMS), linking to participant by unique study identifier only. Data for analysis were extracted from SQL and LIMs as comma delimited files (csv).

2.11 Study team

I was the principal investigator of this study and led the study team. Helen Mangochi was the lead study nurse who assisted with co-ordination of the

team and recruited participants. Two study nurses, Edwin Bullah and Gladys Namacha, recruited patients to the hospital cohort, collected samples and completed follow-up telephone calls. Field workers, Pilirani Kachulu, Gladys Namacha and Suzgo Mkandawire, carried out recruitment, follow-ups and sample collection for the household cohort. James Mango was the study clinical officer, who carried out medical note review and questionnaire completion for patients who had died by the time of enrolment, as well as medical note review to complete direct hospital costing data capture tools. The study laboratory technician, Winnie Bakali, processed stool and urine cultures, carried out AST on blood culture isolates and did POC CRP testing. Rachel Banda and Jacqueline Phiri carried out DNA extractions. The routine blood culture testing and processing of haematology/biochemistry samples was undertaken by MLW core laboratory technicians. Lumbani Makhaza was the study data officer who built the study database and ODK/Teleform data capture tools.

2.12 Ethics and consent

2.12.1 Ethical approvals

Ethical approval for the study was granted by the Malawi College of Medicine Research Ethics Committee (COMREC), protocol number P.10/17/2299 and by the LSTM Research Ethics committee, protocol number 17-063. Sponsorship was provided by LSTM.

2.12.2 Consent and participant remuneration

Informed written consent was obtained from all participants. Potential recruitees were approached by a member of the study team and the study

was explained. Written patient information sheets (PIS) were provided in Chichewa or English. Those willing to participate signed the consent form. If the participant or guardian was unable to read and write, the PIS and consent forms were read to them by a member of the study team, and the consent form signed using a thumbprint, in the presence of an independent witness who was not part of the study team.

If a patient in Cohort 1 lacked capacity to provide consent, consent was sought from the patient's guardian in the presence of an independent witness. If no guardian was available, recruitment was not attempted. If the participant later regained capacity, they were approached for informed written consent and could withdraw if they so wished. All participants could withdraw from the study at any time without giving a reason. Parent or guardian consent was obtained for all children under the age of 18 years. In addition to parental consent, assent was sought from children aged eight years and above, in accordance with WHO guidelines[159]. As consent was not possible for patients who had died by the time of enrolment, all medical records were anonymised by the study clinical officer, who was a Malawi Ministry of Health employee, at the request of COMREC. Once recruited, no other member of the study team had access to any identifiable medical records for these patients. Household participants received remuneration for their time at a rate of 7,500MWK (7.50 GBP/9.21 USD) divided over 4 visits, in line with Malawian guidelines. Hospital participants were not financially remunerated.

2.13 Case definitions

Phenotypic cefoxitin resistance is considered a sensitive screening test for AmpC production[13]. The organisms *Enterobacter* spp., *Serratia marcesens.*, *Citrobacter freundii*, *Providencia stuartii*, *Morganella morganii* and *Hafnia alvei* were therefore defined as 3GC-R if resistant to cefoxitin on AST.

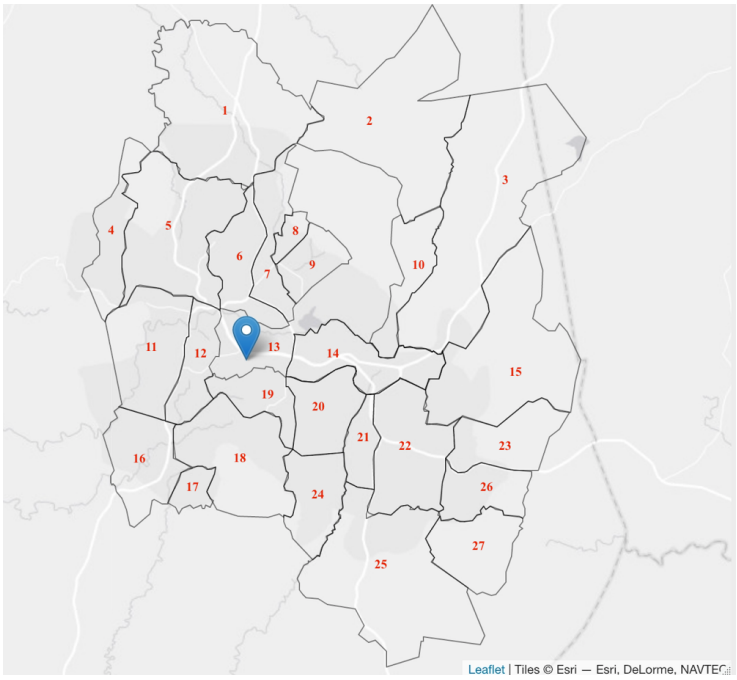
Third-generation cephalosporin resistance in all other organisms was defined as resistance to one or both of cefpodoxime or ceftriaxone on AST testing, regardless of confirmatory ESBL result.

2.14 Appendix

R code used to generate sample sites for cohort 2 using mixture-distribution

```
## Packages
rgdal_1.4-8
raster 3.0-7
lme4 1.1-21

## Draws population density-weighted sample sites

# @param r a raster containing probabilities for sample sites
# @param size number of sample sites
# @param kappa concentration parameter (see details) power to raise
sample probabilities to (0=spatially uniform, 1=proportional to density)
# @param method if 'power' use powered sampling, if 'mixture' use
mixture distribution sampling.
# @details @name samples pixels in a raster with probability proportional
to a function of the pixel value.
#
drawSpatialSamples = function(r, size, kappa,
method=c('power','mixture'))
{
        library(raster)
  pts = rasterToPoints(r, spatial=T)
  names(pts) = c('density')
  tosample = NULL
  if(method[1]=='power') {
          tosample = sample(nrow(pts), size=size,
prob=pts$density^kappa, replace=T)
  }
  else if(method[1]=='mixture') {
        p = pts$density / sum(pts$density)
        p = 1 - (1-p*kappa)*(1-(1-kappa)/nrow(pts))
        tosample = sample(nrow(pts), size=size, prob=p, replace=T)
  }
  else {
        stop("method must be one of c('power','mixture')")
```

```
  }
  plot(r)
  plot(pts[tosample,], pch='+', add=T)
  pts[tosample,]
}

## Load Blantyre City polygon
btPoly = getData("GADM", country="Malawi", level=2)
btPoly = btPoly[btPoly$NAME_2 =="Blantyre City",]

## Load WorldPop raster for Blantyre area, and masks everything
outside the polygon
btWorldPop = raster("~/Desktop/blantyreWorldPop.tif")
btWorldPop = mask(btWorldPop,btPoly)

## Sample points (sample 115 points, compress probabilities 0.9)
sSites = drawSpatialSamples(r=btWorldPop, size=115, kappa=0.9,
method='mixture')
plot(sSites)
```



Appendix Figure 2.1: Blanytre's 27 traditional authority areas, numbered in red with boundaries shown as individual polygons. Blue maker indicates location of QECH. One TA was randomly selected each week for sampling.

1	Michiru
2	South Lunzu
3	Mpanga
4	Chiolomoni
5	Likhubula
6	Nyambadwe
7	Ndirande West
8	Ndirande North
9	Ndirande South
10	Nkolokoti
11	Blantyre West
12	Blantyre Central
13	Blantyre East
14	Chichiri
15	Mzedi
16	Nancholi
17	Chilobwe
18	Soche West
19	Soche East
20	Limbe West
21	Limbe Central
22	Limbe East
23	Bangwe
24	Misesa
25	Bangwe
26	Namiyango
27	Msamba

Chapter 3

Gram negative bloodstream infection in Blantyre, Malawi: a microbiological clinical and epidemiological description

3.1 Overview

In this chapter, I present a microbiological and clinical description of patients with bloodstream Enterobacterales, in Blantyre, and an analysis of risk factors for 3GC-R in these infections. *E. coli* and *K. pneumoniae* were the most common pathogens isolated from pure pathogen cultures in the cohort, at 46.1% and 32.4% respectively. A total of 71.4% of organisms were 3GC-R and 44% were resistant to all readily available antimicrobials. Excluding patients with neonatal sepsis, 62.6% of BSI were healthcare associated. Early-mortality was high, and 31.8% of patients in the cohort had died by the time their blood culture results were available.

Logistic regression was used to identify risk-factors for 3GC-R in BSI. On univariable analysis, there was evidence of an association between prior healthcare exposures and 3GC-R. These exposures were admission to health centres, operations, catheters and ceftriaxone usage prior to the BSI. HIV exposure (defined as positive maternal HIV status in children under 18 months) was associated with 3GC-R, but there was evidence that HIV infection in the older population was negatively associated. Older age was negatively associated with 3GC-R, but with a small effect size. A number of WASH factors were important, and access to piped water for drinking

and sharing a toilet with other households, showed a negative association with 3GC-R.

Multivariable analysis was hindered by missing data and collinearity of variables, and models were likely over-parametrised because of the high number of predictor variables included. Models were generated using multiply imputed data to account for missing-data bias and two methods for collapsing predictor variables (inspection for collinearity and stepwise backward elimination) were attempted. On final multivariable regression, prior operations remained associated with 3GC-R and shared toilets, HIV infection and older age were protective.

3.2 Background and chapter aims

The aims of this chapter are:

1. To describe the characteristics of patients presenting to QECH who have bloodstream infection with Enterobacterales, in terms of demographics, microbiology, clinical features, epidemiology and focus of infection;

2. To identify potential risk factors for BSI with 3GCR-E, with a view to informing empirical antibiotic prescribing protocols at QECH.

3.3 Methods

The general clinical and laboratory methods for the hospital cohort, from which data for this chapter are derived, are described in Chapter 2. A

number of additional specific methods pertaining to this chapter are outlined below.

The additional methods relevant to describing participant characteristics (Aim 1) are described here. The statistical methods relevant to the second aim are briefly outlined in Section 3.3.5, but for clarity are described in detail in the later sections of this chapter, alongside the results.

3.3.1 Focus of infection

As described in Chapter 2, a clinical assessment of each recruited participant was undertaken by me, or by a clinical member of the study team who discussed the patient with me. This assessment was, in part, targeted at determining a likely focus of the patient's BSI (Table 3.1). These categories were derived from existing knowledge of the clinical epidemiology of Gram-negative BSI in other settings[160]. An additional category, gut translocation and non-focal, was specified to reflect BSI caused by the non-physiological passage of gastrointestinal microflora through a friable intestinal epithelial barrier into blood and extra-intestinal sites[161]. This phenomenon is well-described in HIV and other immune compromised hosts[161, 162] and is therefore likely to be important in this high HIV prevalence setting.

Following review of clinical history, examination and available investigations, participants were classified into likely rather than definite focus, because of a desire not to overclassify non-focal sepsis in a setting where lack of diagnostic resources frequently limit the ability to definitively

confirm focus of infection. Frequency of each category is presented, stratified by age-group.

Table 3.1: Categories for clinically suspected focus of infection.

Focus
Non-focal
Gut translocation and non-focal
Post-operative and non-focal
Clear focus of infection
Urinary tract infection (catheter associated)
Urinary tract infection (non-catheter associated)
CNS
Skin and soft tissue
Gastrointestinal (hepatobiliary)
Gastrointestinal (non-hepatobiliary)
Cardiovascular system
Respiratory tract infection (other than VAP)
Reproductive tract infection
Surgical site infection
Bone and joint
VAP
Other
Unknown

Note:
CNS = Central nervous system, VAP = Ventilator associated pneumonia

3.3.2 Epidemiological attribution

All BSI were categorised according to their suspected onset (community or hospital) and their relationship to healthcare (healthcare- or non-healthcare associated). These are defined in Table 3.2 [163]. Relative proportions of BSI in each category are presented. A clinical assessment of each patient was again made in order to accurately determine the appropriate category.

Table 3.2: Definitions of epidemiological attribution categories. Adapted from[163] and [164].

Category	Definition
Community onset, healthcare-associated	BSI which develops in the community, in a patient who has received healthcare in either the community or hospital in the previous 28 days
Community onset, non-healthcare-associated	BSI which develops in the community, in a patient who has not received healthcare in either the community or hospital in the previous 28 days
Hospital onset	BSI which develops in a patient 48 hours or more after admission to hospital (either QECH or a hospital from which the patient was directly transferred to QECH, without spending a night at home)
Early-onset neonatal sepsis	Bloodstream infection occurring in young infants in the first week of life
Late-onset neonatal sepsis	Bloodstream infection occurring in young infants from one week to 3 months of life

Note:
BSI = bloodstream infection, QECH = Queen Elizabeth Central Hospital

3.3.3 Anthropometry

A nutritional assessment was carried out on participants aged 0-59 months in accordance with WHO standards[165]. These children had their weight, height and mid-upper arm circumference (MUAC) measured. Results for each anthropometric index are presented as medians, and weight-for-height z-scores were calculated against WHO Growth Reference Standards[165], using the *zscorer* package in R[166].

3.3.4 Household wealth index

A composite household wealth variable was constructed using the proxy means test method[167], based on the 1998 Malawi Integrated Household Survey[168]. This linear model incorporates eight urban household variables which have been shown in Malawi to correlate with overall household consumption level, allowing participants' household wealth to be ranked based on their ability to meet household consumption needs[167, 169]. The coefficients used in this model are shown in Table 3.3. Using this model, a continuous variable for wealth was generated by multiplying the identified variables by their respective co-efficient, such that higher values represent wealthier households. The resulting composite wealth variable was used as a covariate in a univariable logistic regression to assess its association with 3GC-R and 3GC-S outcomes as described in Section 3.3.5.

Table 3.3: Coefficients for urban proxy means test model. Adapted from[167].

Response variable: log household welfare indicator	
Predictor variables	Co-efficient
Household owns fridge	0.518
Household size	-0.306
Household size squared	0.016
Age of head of household	0.005
Education level of head of household	0.151
No. of salaried household members	0.061
Household owns a motor vehicle	0.704
Household gets lighting from electricity or gas	0.280
Household owns a bed	0.247
Blantyre City	-0.037
Constant	2.347

3.3.5 Statistical methods

Demographics, pre-hospital healthcare exposures, clinical features and household level covariates are presented in tables as medians and interquartile ranges (in the case of continuous variables) or as proportions with exact binomial 95% confidence intervals[170] (in the case of categorical variables). These variables have been selected a priori as potentially expected to be associated with 3GC-R infection and are presented in tables stratified by the 3GC susceptibility status. Univariable associations of 3GC susceptibility status were assessed from univariable logistic regression models with each variable considered alone as a predictor, using the *glm* function in R. P-values are presented in the relevant tables.

To adjust for confounding and produce unbiased risk-factor estimates, predictor variables for which $p<0.20$ on univariable analysis, were selected for inclusion in a multivariable logistic regression, with the aim of further refining the identification of risk factors for 3GC-R in BSI. Odds ratios (OR) with 95% confidence intervals (95% CI) were calculated and are presented in the text and in summary tables.

There were a number of challenges associated with building this model and so further details of the process of model evolution are provided in the relevant results sections of this chapter. All statistical analyses were conducted in R.studio.

3.4 Results

3.4.1 Study population

Between January 31st 2018 and January 13th 2020, 728 patients with a relevant blood culture result were screened by the study team and 346 participants were enrolled (Figure 3.1).

During the same time period, 1211 blood cultures done by the core MLW laboratory were positive for the Gram-negative pathogens of interest, meaning that overall, 60.1% (728/1211) of all relevant cultures were screened and 28.6% (344/1211) were recruited. Reasons that not all blood cultures were screened include gaps over weekends and holidays when the blood culture bench and LIMS were not reviewed and delayed laboratory reporting of final results. It is also likely that not all screened patients were recorded on screening logs. The two most common reasons screened patients were not recruited were patient choice and inability of study team to locate the patient (Figure 3.1).

Ten participants (2.9%) were excluded from the study post-recruitment and not included in the analysis. Of these, two were withdrawn by the study team because of errors in the core laboratory which meant the final organism identification changed to *Salmonella*, shortly after the participant had been recruited. Two participants absconded from hospital with no known in-hospital outcome and were therefore withdrawn. The remaining six participants withdrew from the study at their own request.

A further two participants chose to withdraw from the study during the 180-day follow-up period, but were happy for their baseline and enrolment data to be used, therefore were included in the analysis of primary outcomes (Figure 3.1).

Of the 336 participants included in the primary analysis, 229 (68.2%) were alive at the point of recruitment and were enrolled using prospective data and sample collection. The remaining 107/336 (31.8%) had died by the time of recruitment and were enrolled using retrospective note review. These latter participants did not have blood, stool or urine samples available for analysis.

10/229 alive participants were lost to follow-up by 28 days so were not included in post-primary analysis. A further 11 participants were lost to follow-up by 90- days and a further 10 participants were lost to follow-up by 180 days.

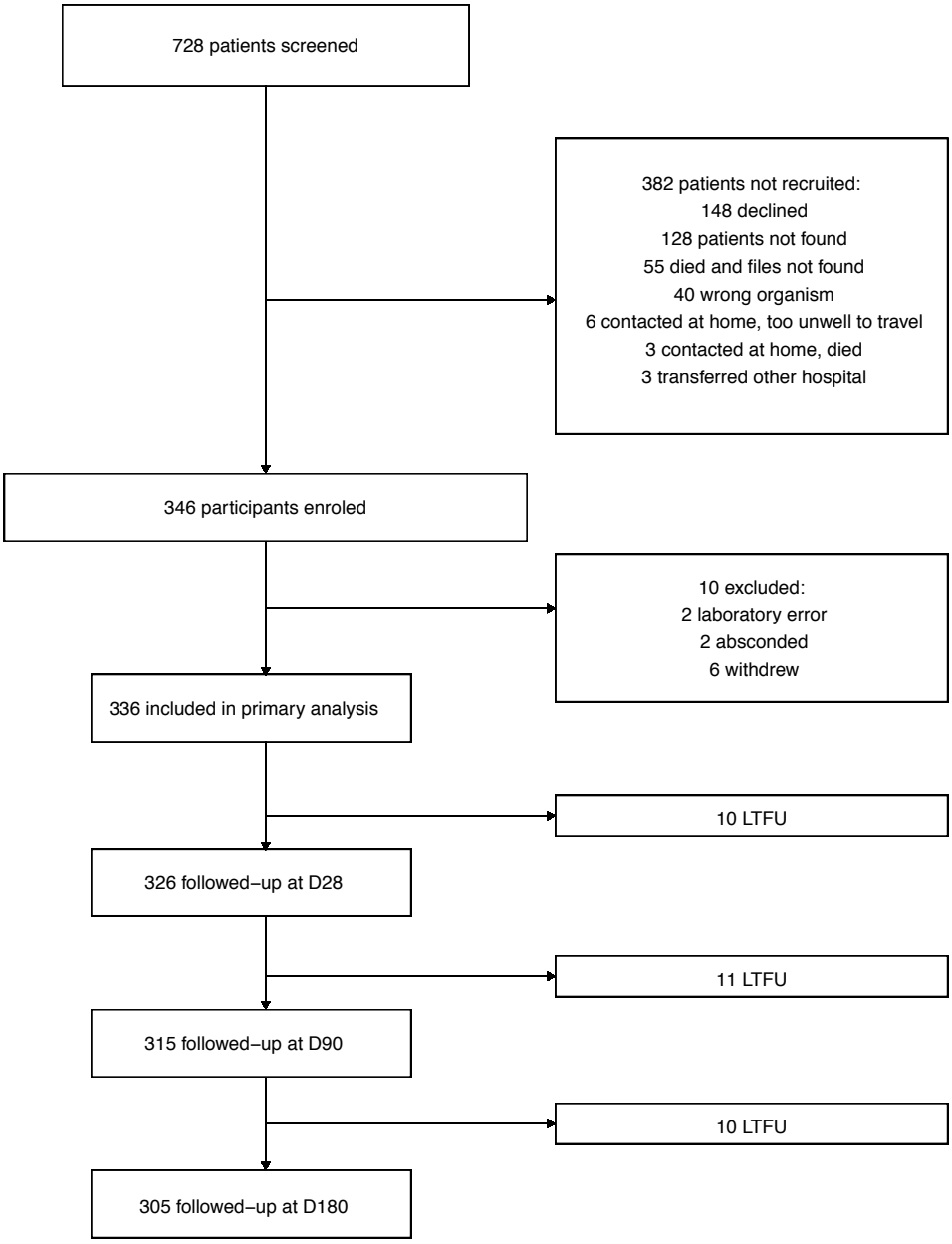


Figure 3.1: Study population at recruitment and follow up. LTFU = lost to follow-up.

3.4.2 Timing of sampling and recruitment

The median time from admission to blood culture sampling amongst participants who were admitted was 3 days (IQR 0-7) and the median time from blood culture sampling to recruitment was 5 days (IQR 4-7).

Forty participants were not admitted to QECH following their blood culture sampling, so are not included in these figures.

The overall median time from admission to recruitment for participants in the cohort was therefore 8 days (IQR 5-13). This is of relevance when considering the timing of blood tests and urine samples, which were therefore often being done over a week into the patient's admission. There was no difference in these time periods between participants with 3GC-R or 3GC-S infection (Figure 3.2).

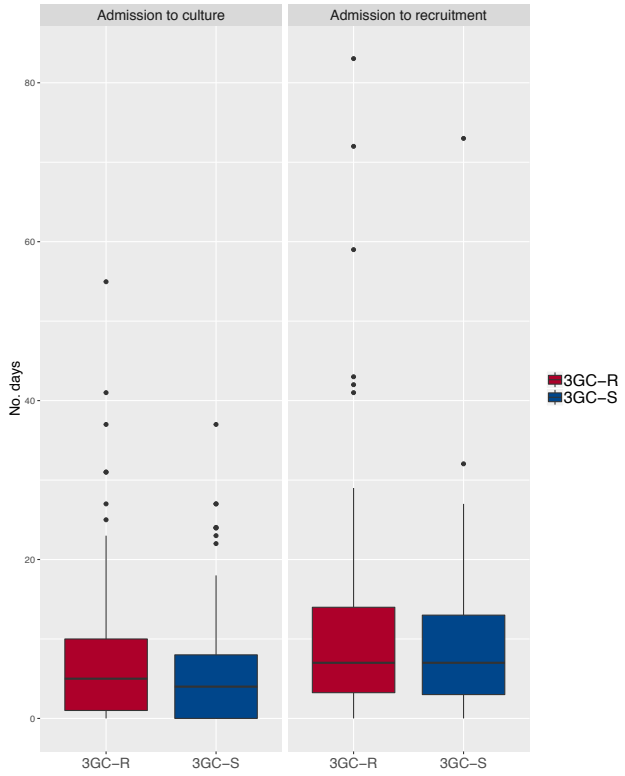


Figure 3.2: Sample and recruitment timing, stratified by 3GC susceptibility status. Admission to culture = time from QECH admission to blood culture sampling. Admission to recruitment = time from QECH admission to participant recruitment.

3.4.3 Microbiology

3.4.3.1 Overview

E. coli and *Klebsiella pneumoniae* were the most common pathogens isolated from pure pathogen cultures in the cohort, representing 46.1% (147/319) and 34.2% (109/319) of pure growth Gram negative BSI respectively. A further 17 blood cultures grew mixed Gram negative organisms:

Six *K. pneumoniae*/*Acinetobacter* spp.;

One *Acinetobacter* spp./*Aeromonas* spp.;

Two *K. pneumoniae/Enterobacter cloacae*;

One *E. coli*/*Acinetobacter* spp.;

Three *E. coli*/*K. pneumoniae*;

One *Acinetobacter* spp./*Pantoea* spp.;

Two *E. coli*/*Pantoea* spp.;

One *Acinetobacter* spp./*Serratia* spp.

Including mixed growth cultures, a total of 252/353 (71.4%) organisms were 3GC-R. Most K. pneumoniae were 3GC-R (110/123 [89.4%]) and half of *E. coli* were 3GC-R (80/157 [50.9%]). Figure 3.3 shows 3GC susceptibility by species.

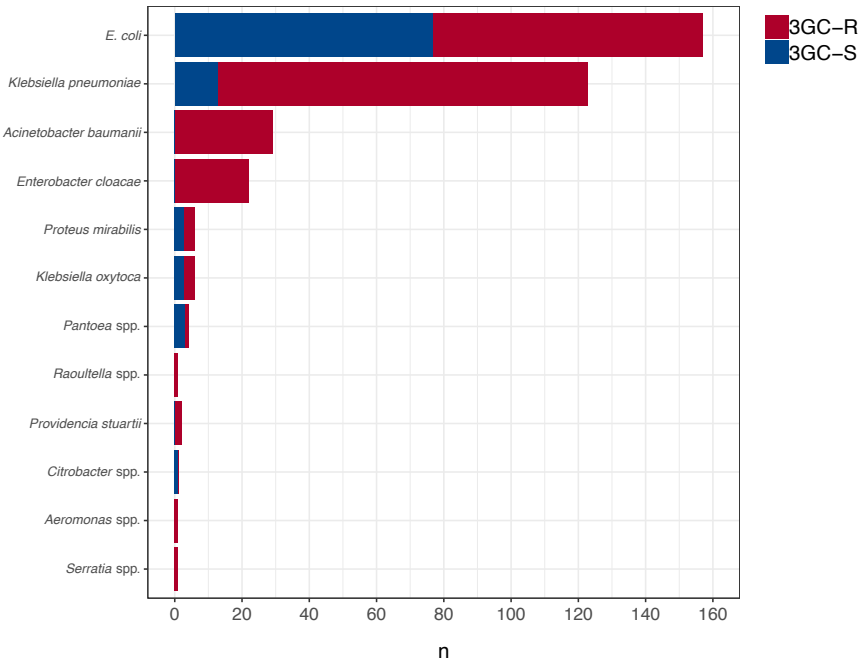


Figure 3.3: Third-generation cephalosporin susceptibility status for all organisms, stratified by species.

All but five 3GC-R organisms were ESBL-producing on combination disc testing. The five organisms which were not found to be ESBL-producers by this method, were tested a second time against the 3GC indicator discs and all were confirmed as 3GC-R.

K. pneumoniae was the most commonly isolated pathogen amongst young infants (aged 0-3 months), where it represented 63% (63/100) total pure BSI (Figure 3.4). Overall, *K. pneumoniae* tended to affect younger age groups (median age of 1.9 months [IQR 0.5 months to 7.2 years]). *E. coli* on the other hand, was more evenly spread across age groups (Figure 3.4) and affected an older median age of 36.1 years (IQR 3.0-54.2 years).

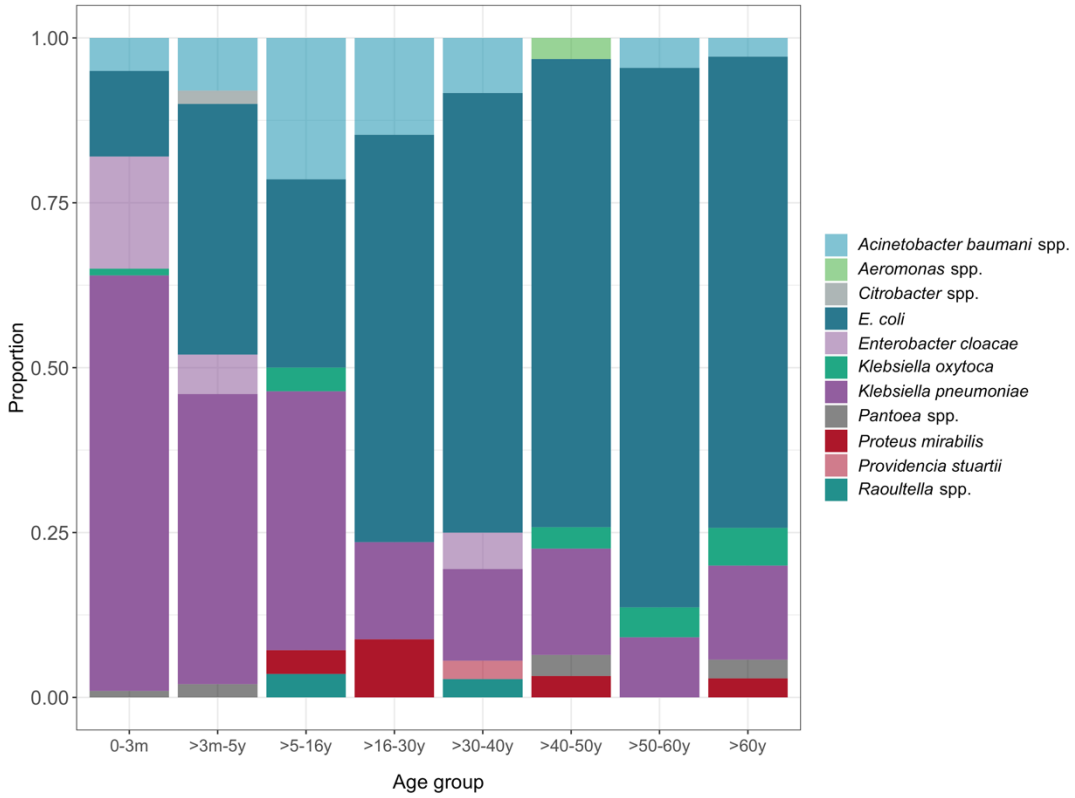


Figure 3.4: Blood culture results, stratified by age-group of participants. m=months, y=years.

3.4.3.2 Resistance to readily available antimicrobials

At QECH, readily available antimicrobials are ciprofloxacin (oral, PO), ceftriaxone (intravenous, IV), amoxicillin (PO), co-trimoxazole (PO) and gentamicin (IV), the route of administration is important as the lack of IV ciprofloxacin precludes the use of this agent in infants. Chloramphenicol and co-amoxiclav are occasionally available, though the former is relatively contraindicated in neonates and the latter is not recommended for treatment of invasive infection with ESBLs. Meropenem, amikacin and piperacillin-tazobactam are not routinely available, however they are sporadically available by donation, especially to the department of paediatrics.

The results of phenotypic AST for all tested organisms are shown in Figure 3.5 and are shown broken down by organism and antimicrobial in Figure 3.6. At the time of writing, of the 17 mixed cultures, only one from each pair had undergone full AST and so this figure does not include the second organism in the case of mixed culture.

Ciprofloxacin resistance was detected in 71.2% (169/236) 3GC-R organisms and in 15.0% (15/100) 3GC-S organisms overall. 91.9% (68/74) of 3GC-R *E. coli* and 61.0% (64/105) of 3GC-R *K. pneumoniae* were resistant to ciprofloxacin.

Chloramphenicol resistance was detected in 48.5% (103/212) of all 3GC-R Enterobacterales, but not tested in *Acinetobacter* and in 20/100 (20%) 3GC-S organisms overall. 28.4% (21/74) of 3GC-R *E. coli* and 61.9% (65/105) of 3GC-R *K. pneumoniae* were resistant to chloramphenicol.

Gentamicin resistance was detected in 83.4% (197/236) of all 3GC-R organisms and in 39/100 (39.0%) 3GC-S organisms overall. 78.4% (58/74) of 3GC-R *E. coli* and 96.2% (101/105) of 3GC-R *K. pneumoniae* were resistant to gentamicin. Almost all, 291/312 (93.2%) of organisms tested were resistant to ampicillin and to co-trimoxazole 305/336 (91.0%).

Overall, 44.6% (150/336) of organisms were resistant to all five readily available antibiotics and hence locally untreatable, although this figure rises in neonates when "unusable" antibiotics are considered. This was 62/118 (52.5%) in K. pneumoniae and 57/151 (37.7%) in *E. coli*.

There was no phenotypic meropenem resistance detected amongst the tested Enterobacterales, though two isolates of *E. coli* had MICs close to the zone diameter breakpoint (16mm with resistance defined as <16mm). Six isolates of *Acinetobacter* spp. displayed meropenem resistance (6/24 [25%]) but a likely mechanism for this, is permeability defect caused by porin loss. Only 8/118 (6.8%) of *K. pneumoniae* and 3.3% (5/151) *E. coli* displayed amikacin resistance.

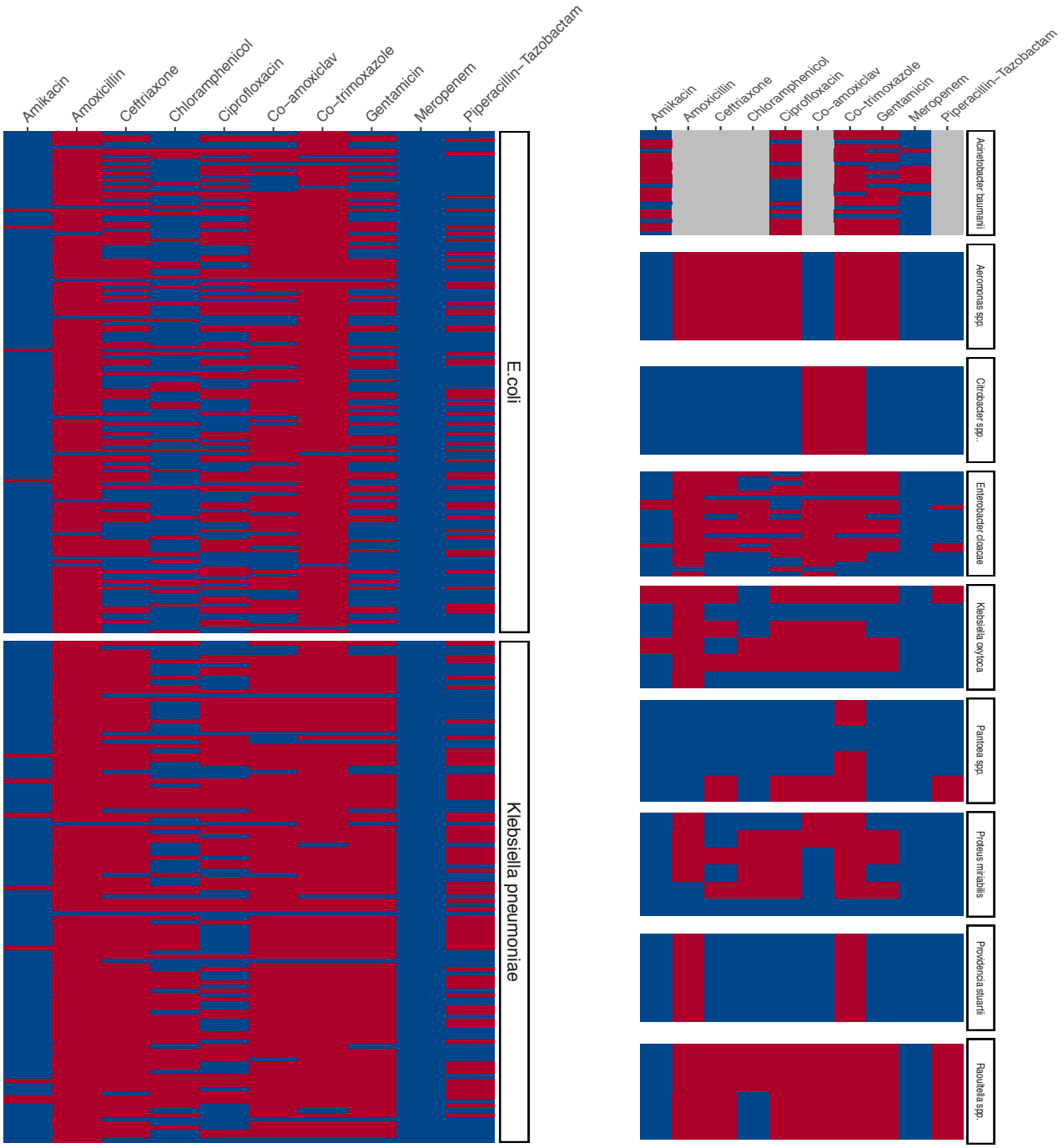


Figure 3.5: Phenotypic antimicrobial resistance profiles for all tested isolates using EUCAST breakpoints. Each row is one isolate and each column is the tested antimicrobial, Red = resistant, blue = sensitive. Grey = not tested. All isolates of *Acinetobacter* were 3GC-R under BSAC breakpoints, but there are no EUCAST breakpoints for *Acinetobacter*/3GC so these are not shown here.

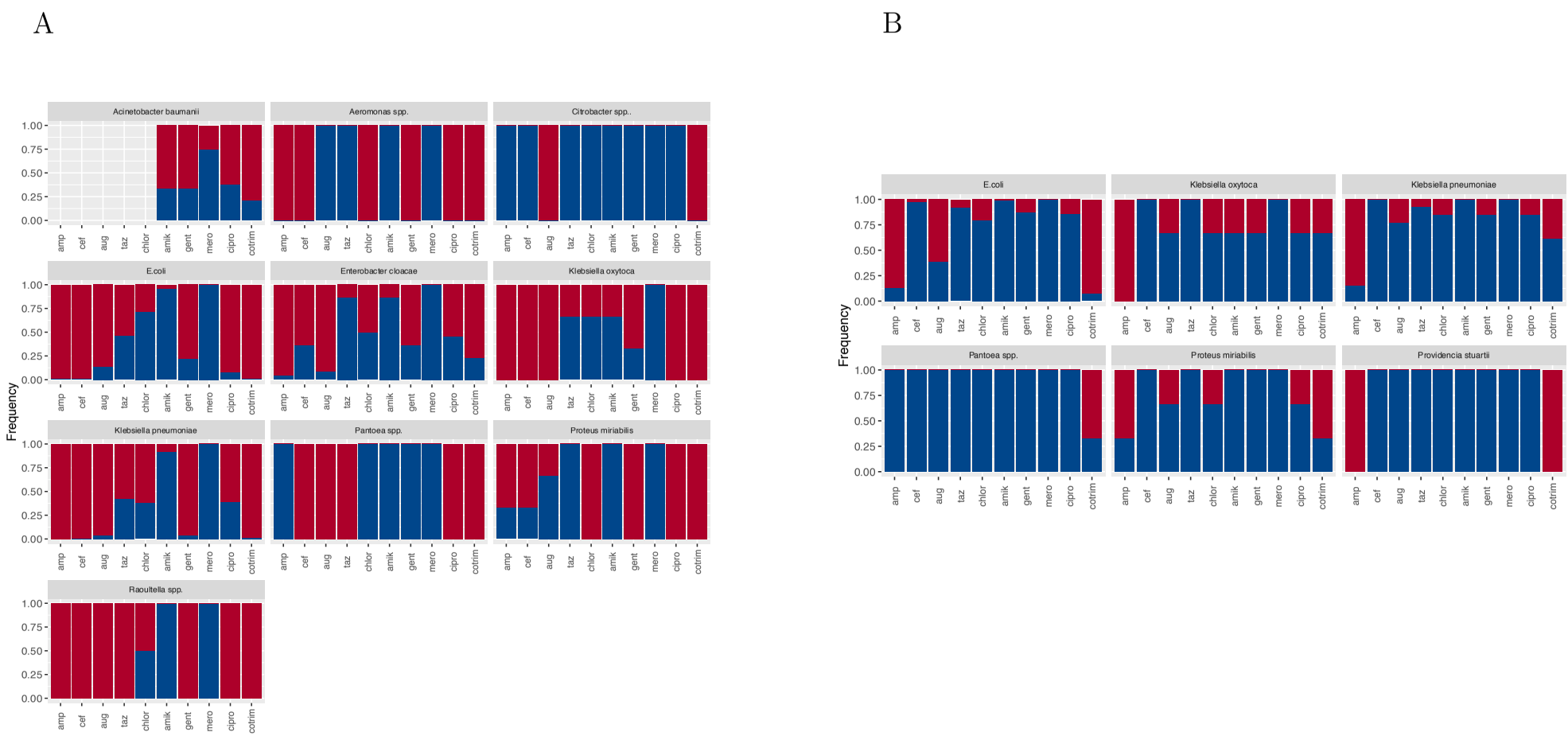


Figure 3.6: Proportions of isolates resistant to each tested antimicrobial in (A) 3GC-R organisms and (B) 3GC-S organisms. Note that one *E. coli* and one *K. pneumoniae* were sensitive to ceftriaxone (hence shown as blue in panel (A) but resistant to cefpodoxime and confirmed as ESBL producers on combination disc testing. Amp=ampicillin, cef = ceftriaxone, aug = co-amoxiclav, taz = piperacillin-tazobacatam, chlor = chloramphenicol, amik = amikacin, gent = gentamicin, mero = meropenem, cipro = ciprofloxacin, cotrim = co-trimoxazole.

3.4.4 Participant characteristics

Table 3.4 shows the baseline characteristic for all recruited participants, stratified by age group. Participants were young, with an overall median age of 12.1 years (IQR 1.0 month – 41.4 years) and a median age amongst adults of 42.8 years (IQR 32.5- 58.0). There were 100 young infants, (aged under 3 months) and of these, 75 were neonates (aged under 28 days). Figure 3.7 shows the overall age distribution of the cohort.

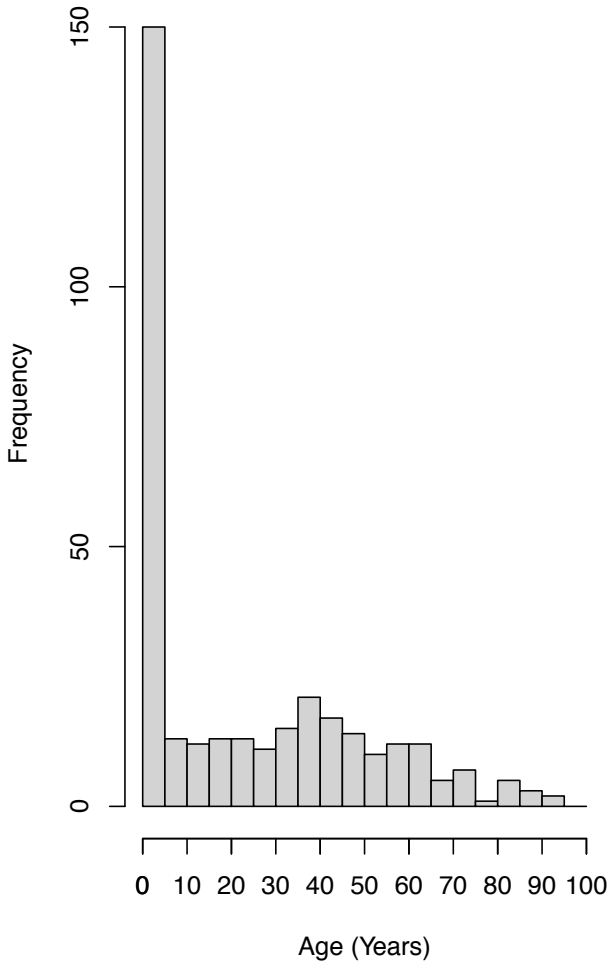


Figure 3.7: Age distribution of recruited participants

Table 3.4: Participant characteristics.

Variable	Adults ≥ 16 years n=159	Children < 16 years n=177
Demographics		
Age	42.8y (32.5y-58.0y)	1.6m (0.5m-20.6m)
Female sex	82/159 (51.6 [43.5-59.6])	96/177 (54.2 [46.6-61.7])
Male sex	77/159 (48.4 [40.4-56.5])	81/177 (45.8 [38.3-53.4])
HIV status		18 months-15years:
HIV positive	91/155 (58.7 [50.5-66.5])	6/43 (14.0 [5.3-27.9])
HIV negative	64/155 (41.3 [33.4-49.5])	37/43 (86.0 [72.1-94.7])
HIV unknown	4/155 (2.6 [0.7-6.5])	7/50 (14.0 [5.8-26.7])
<18 months old:		<18months:
HIV positive	-	3/119 (1.5 [0.3-4.3])
HIV exposed/unknown status	-	19/119 (9.5 [5.8-14.5]))
HIV unexposed	-	97/119 (48.7 [41.6-55.9])
Unknown exposure/infection	-	8/127 (6.3 [27.6-12.0])
ART status		
Current ART	74/91 (81.3 [71.8-88.7])	5/9 (55.6 [21.2-86.3])
Months on ART	53.5 (13.5-106.6)	14.2(10.3-111.5)
ART regimen:		
TDF/3TC/EFV	56/74 (75.7 [64.3-84.9])	1/5 (20.0 [0.5-71.6])
Other regimen	10/74 (13.5 [6.7-23.5])	4/5 (80.0 [28.4-99.5])
Current CPT	68/91 (74.7 [64.5-83.2])	6/9 (75.0 [34.9-96.8])
Education		
No formal schooling	3/114* (2.6 [0.5-7.5])	-
Any primary education	57/114 (50.0 [40.5-59.5])	-
Any secondary	34/114 (29.8 [21.6-39.1])	-
College or higher level	21/114 (18.4 [11.8-26.8])	-
School age and currently in education	-	22/30 (73.3 [54.1-87.7])
Employment		
Unemployed	61/115 (53.0 [43.5-62.4])	-
Paid employee	19/115 (16.5 [10.2-24.6])	-
Self-employed	29/115 (25.2 [17.6-34.2])	-
Student	6/115 (5.2 [1.9-11.0])	-

Note:
Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR.
ART = Antiretroviral therapy, TDF = Tenofovir, 3TC = Lamivudine, EFV = Efavirenz, CPT = Cotrimoxazole preventive therapy

Over half of adults were known to be HIV infected (91/155 [58.7%]). In children aged between 18 months and 16 years, 14.0% (6/43), were known

to be HIV infected. In children under 18 months old, 1.5% (3/119), were HIV infected and 16.0% (19/119), were HIV exposed (defined as maternal HIV infected). These latter HIV exposed children did not have HIV viral load results available, so their infection status was not confirmed during the course of this study. Most participants with HIV infection were taking ART (79/100 [79.0%]) and 74% (74/100) were taking cotrimoxazole preventive therapy (CPT). Median duration of ART therapy was 26 months.

Table 3.5 shows key baseline characteristics stratified by 3GC susceptibility status of the BSI (3GC-R vs 3GC-S). There was some evidence of a negative association of age on 3GC-R (OR 0.98 95% CI 0.97-0.99), though with a very small effect size. HIV appeared to be negatively associated with 3GC-R in participants over 18 months of age (OR 0.33 95% CI 0.18-0.58), but HIV exposure was positively correlated (OR 2.50 95% CI 0.65-16.4).

Table 3.5: Demographics and HIV status, stratified by 3GC susceptibility status.

Variable	3GC-R	3GC-S	p
Demographics			
Age, Median	**2.8y (1.0m-36.3y)**	**33.8y (2.9y-55.0m)**	**<0.001**
Female sex	121/236 (48.7 [42.2-55.3])	48/100 (48.0 [37.9-58.2])	0.903
Male sex	115/236 (51.3 [44.7-55.7])	52/100 (52.0 [41.8-62.1])	0.903
HIV status			
>18 months old			
HIV positive	**48/124 (38.7 [30.1-47.9])**	**49/74 (66.2 [54.2-76.8])**	**<0.001**
HIV negative	**76/124 (61.3 [52.3-69.9])**	**25/74(33.7 [23.1-45.7])**	**<0.001**
HIV unknown	8/132 (6.0 [2.7-11.6])	3/77 (3.9 [0.8-11.0])	0.918
<18 months old			
HIV exposed	**20/96 (24.0 [15.8-33.7])**	**2/23 (8.7 [1.1-28.0])**	**0.193**
HIV unexposed	76/96 (79.2 [69.7-86.8])	21/23 (91.3 [72.0-98.9])	0.366
Unknown exposure	8/104 (7.6 [3.3-14.5])	0	-
Current ART (if HIV infected)	**37/48 (77.1 [62.7-88.0])**	**42/49 (85.7 [72.8-94.1])**	**<0.01**
Current CPT (if HIV infected)	38/48 (79.2 [65.0-89.5])	36/49 (73.5 [58.9-85.0])	0.506

Note:
Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR.
Variables for which p<0.2 on univariate analysis, which will be analysed in multivariate model, are shown in bold
ART = Antiretroviral therapy, CPT = Cotrimoxazole preventive therapy.

3.4.5 Clinical characteristics, physiology and laboratory investigations

Figure 3.8 shows the wide range of primary admission diagnoses for participants in the cohort. These were extracted from the medical records and for the most part represent clinical diagnoses, which are the norm at QECH where diagnostics are limited. Elective or emergent paediatric surgery was the most frequent reason for admission, accounting for 19.3% (65/336) of all participant diagnoses. Of these surgical cases, most were gastrointestinal in nature, with Hirschprung's disease, gastroschisis and bowel fistulae accounting for 70.8% (46/65).

Sepsis without a stated cause was common, representing 14.5% (49/336) of all diagnoses and malignancy of various types (haematological, oesophageal, cervical, Wilm's tumor and Kaposi's sarcoma) was the primary diagnosis in 8.6% (29/336) of admissions. TB accounted for 4.5% (15/336) admission diagnosis, but again, this was a suspected and not confirmed diagnosis.

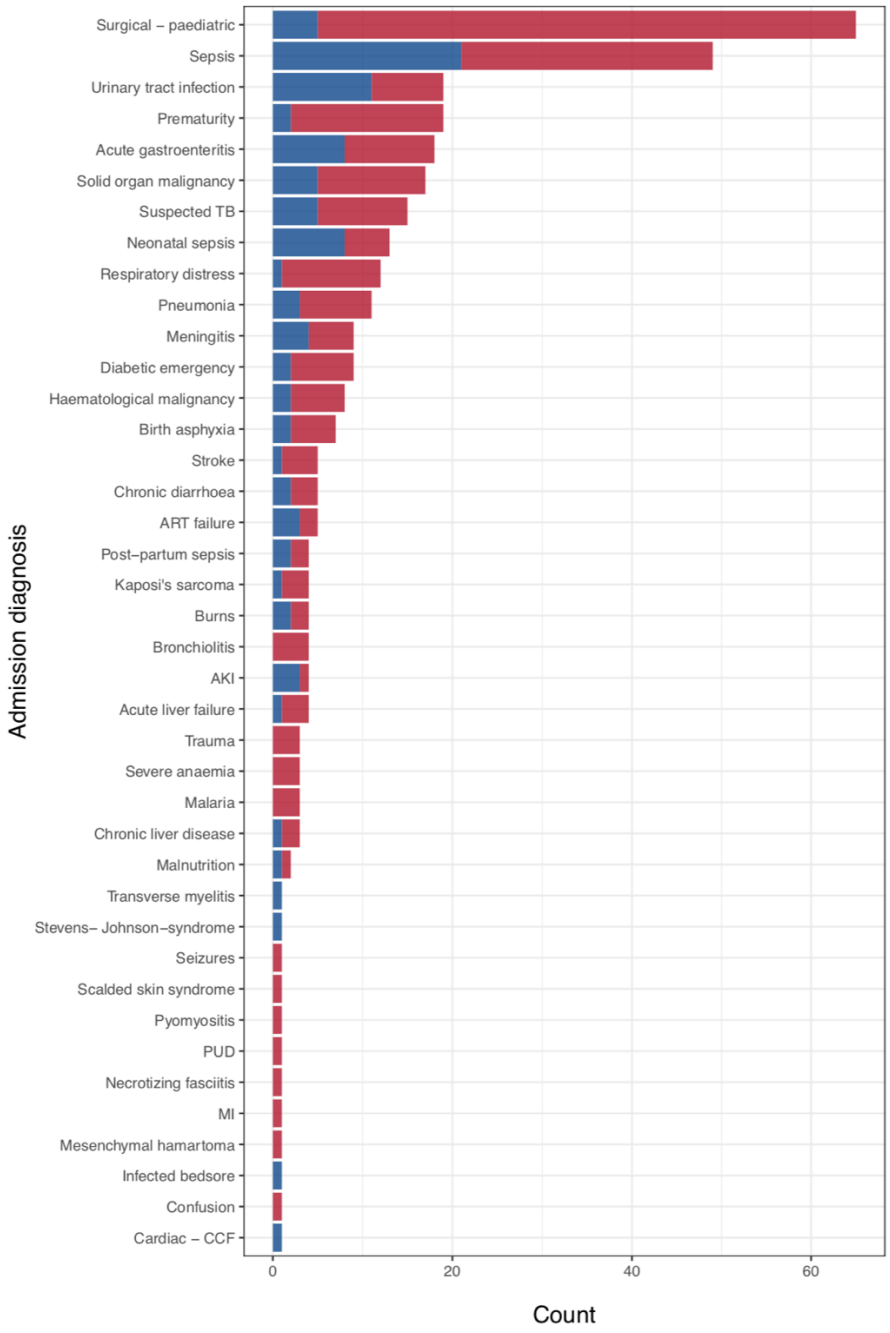


Figure 3.8: Primary admission diagnosis for all 336 participants, stratified by 3GC-R susceptibility status. Red = 3GC-R, blue = 3GC-S.

Figure 3.9 shows the organisms implicated in the 12 most common admission diagnoses. *E. coli* was the most common organism implicated in patients with suspected urinary tract infection (UTI), adults with suspected sepsis, solid organ malignancy and gastroenteritis.

K. pneumoniae was the most common organism implicated in paediatric surgical admissions and in neonatal pathologies (prematurity, respiratory distress and neonatal sepsis).

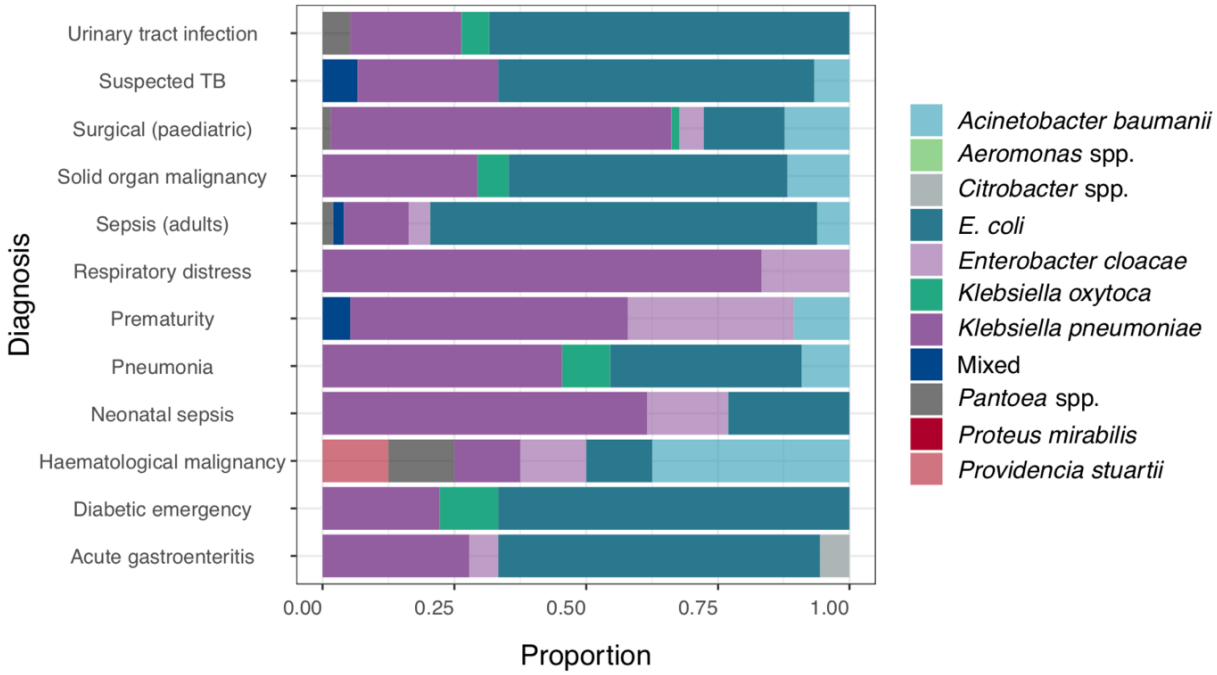


Figure 3.9: Organisms implicated in the twelve most common admission diagnoses, shown as proportions for each diagnosis. Mixed = mixed Enterobacterales.

Table 3.6 shows participants' vital signs and laboratory investigations stratified by 3GC-R status. In general, the adult cohort appeared to have less severe sepsis based on conventional physiological parameters, than might be expected. Median temperatures were $< 38.0^{o}$C, median systolic blood pressure in adults $>$ 100mmHg and median respiratory rates only slightly raised (22-24 min^{-1}). Pooled vital sign parameters are difficult to present for children, on account of the different normal physiological ranges in different age groups.

There was some evidence, however, that temperature on admission was lower in children with 3CG-R BSI than in children with 3GC-S BSI, potentially reflective of being more unwell and/or more immunosuppressed, but with OR crossing 1.0 (OR 0.77 95% CI 0.58-1.0).

Using vital signs that are applicable to the majority of age groups, the strongest evidence that patients with 3GC-R might be sicker than those with 3GC-S BSI, derives from the unable to stand variable. Overall, inability to stand at the time of admission (if normally able to stand) was associated with 3GC-R (OR 2.21 95% CI 1.24-3.92), as well as at the time of recruitment (OR 2.65 95% CI 1.48-4.77). There was also some evidence that admission temperature was lower on admission for patients of all ages with 3GC-R (OR 0.83 95% CI 0.70-0.97).

Despite high reported ART coverage, the overall median CD4 count was < 200 cells μL^{-1} in adults and there were no significant associations of any blood parameters with 3GC susceptibility (Table 3.6). Over half of participants who reported to be on ART for greater than 6 months 31/57 (54.4%) had a CD4 count of less than 100 cells μL^{-1} therefore fulfilling WHO criteria for immunological HIV treatment failure.

Table 3.6: Participant physiology and blood parameters, stratified by 3GC susceptibility status.

Variable	3CC-R	3GC-S	p
Admission physiology (adults)†			
Temperature (°C)	37.3 (36.4-38.3)	37.6 (36.4-38.6)	0.778
Heart rate (min^{-1})	112 (93 -123)	110 (98-128)	0.819
Systolic BP (mmHg)	109 (94-132)	108 (88-131)	0.707
Diastolic BP (mmHg)	70 (60-84)	66 (56-85)	0.366
Respiratory rate (min^{-1})	24 (20-28)	22 (19-26)	0.199
Recruitment physiology (adults)† †			
Temperature (°C)	36.3 (36.1-37.2)	36.5 (36.1-37.2)	0.405
Heart rate (min^{-1})	96 (84-113)	99 (86-111)	0.865
Systolic BP (mmHg)	118 (99-123)	106 (96-128)	0.436
Diastolic BP (mmHg)	74 (67-80)	70 (61-80)	0.534
Respiratory rate (min^{-1})	22 (20-24)	22 (18-26)	0.714
Physiology (children)			
Temperature on admission (°C)	36.8 (36.1-37.8)	37.1 (36.5-38.6)	0.056
Temperature on recruitment (°C)	36.5 (36.1-36.9)	36.7 (36.3-37.1)	0.447
Physiology (all ages)			
Temperature (°C) admission	37.1 (36.2-38.2)	37.4 (36.4-38.3)	0.023
Temperature (°C) recruitment	36.3 (36.1-36.7)	36.5 (36.1-37.2)	0.405
Unable to stand on admission*	117/158 (74.1 [66.5-80.7])	44/78 (56.4 [44.7-67.6])	0.007
Unable to stand on recruitment*	109/231 (47.2 [40.6-53.8])	38/74 (51.4 [39.4-63.1[)	0.001
Blood parameters (all ages)††			
White cell count (x109 L^{-1})	9.7 (6.8-14.3)	10.1 (5.7-13.9)	0.493
Haemoglobin (x109 dL^{-1})	9.5 (7.8-11.2)	9.4 (7.4-11	0.952
Platelet count (x109 L^{-1})	212(109-354)	250 (139-395)	0.745
Creatinine (mmol L^{-1})	84 (47-102)	65 (40-86)	0.209
CRP (mg L^{-1})	62(27-119)	107(42-120)	0.673
Lactate (mmol L^{-1})	2.9 (2.2-3.2)	3.1(2.5-3.2)	0.992
CD4 (cells μL^{-1}) (adults)	127 (22-359)	165 (38-293)	0.800
CD4 (cells μL^{-1}) (children)	1349 (690-1975)	no data	-

Note:
Adults refers to participants at least 16 years of age. Children refers to participants younger than 16 years.
† Admission physiology was as recorded as part of routine hospital care therefore has missing data for heart rate (n=22), respiratory rate (n=99), blood pressure (n=161)
†† Recruitment physiology and blood tests only available for participants alive on recruitment and are therefore missing for n = 107
*Applies only to those participants normally able to stand
Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR. Variables for which $p<0.2$ on univariable analysis, are shown in bold
BP = Blood pressure

3.4.6 Epidemiological attribution

Excluding neonatal sepsis, 50.5% (142/281) of BSI episodes were classified as hospital onset, 12.1% (34/281) were classified as community-onset, healthcare associated and 37.4% (105/281) were community-onset, non-healthcare associated. Therefore in these patients, 62.6% (176/281) of BSI were associated with healthcare. Amongst neonatal sepsis episodes, 94.5% (52/55) had early-onset sepsis and the remaining 5.5% (3/55) had late onset-sepsis. Of these neonates, 45.5% (25/55) were born at QECH, 47.3% (26/55) were born in other hospitals or health-centres and the remaining 9.1% (4/55) were born at home. Thus a total of 30/55 (54.5%) neonates with neonatal sepsis were "outborn".

Figure 3.10 shows proportions of each epidemiological category stratified by organism. *E. coli* were mostly community onset non-healthcare associated, 69/111 (62.2%). By contrast, 70/115 (60.9%) *K. pneumoniae* were healthcare associated or hospital-onset.

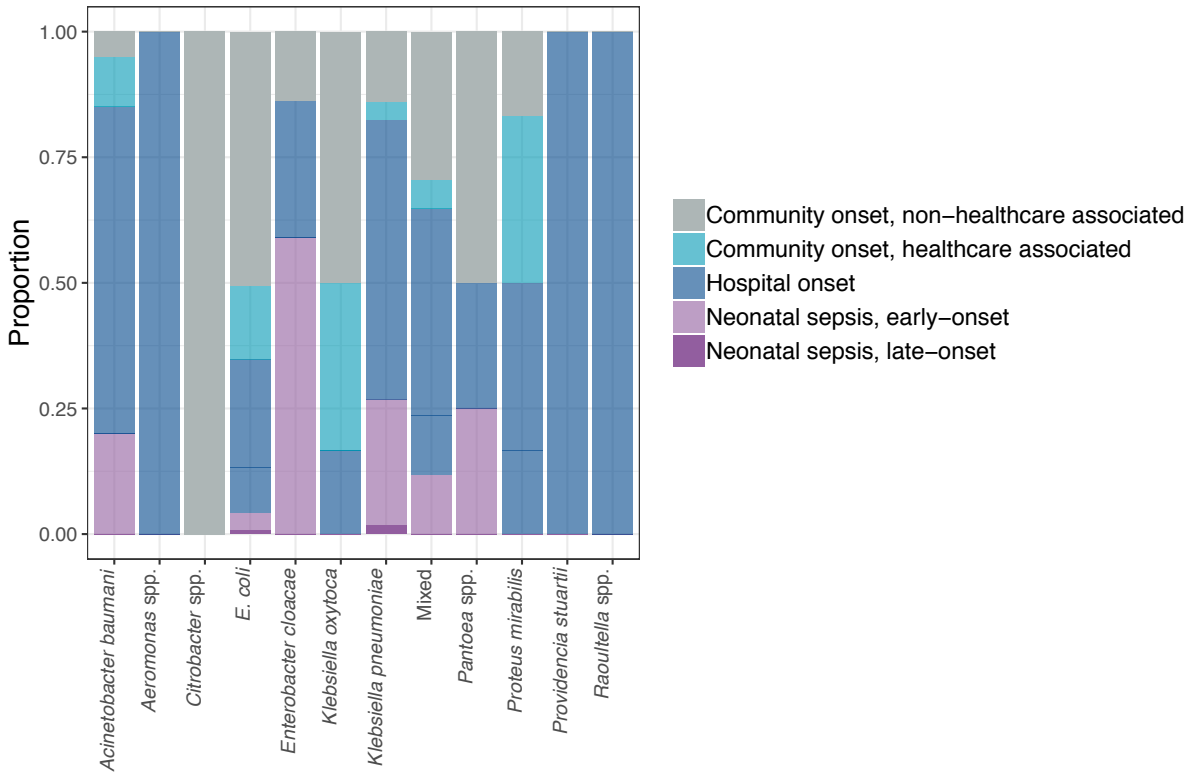


Figure 3.10: Epidemiological attribution, stratified by organism

Table 3.7 stratifies these categories by 3GC susceptibility. A significantly higher proportion of hospital onset BSI, were 3GC-R than 3GC-S (36.0% vs 21.0%, p<0.001). By contrast, a significantly higher proportion of non-healthcare associated BSI were 3GC-S than 3GC-R (58.0% vs 14.0%, p<0.001).

Table 3.7: Epidemiological attribution, stratified by 3GC susceptibility status.

Category	3GC-R	3GC-S	p
Hospital onset	121/236 (36.0 [30.9-41.4])	21/100 (21.0 [12.5-30.3])	<0.001
Community onset, healthcare associated	24/236 (7.1 [4.6-10.4])	10/100 (10.0 [4.9-17.6])	0.999
Community onset, non-healthcare associated	47/236 (14.0 [10.5-18.2])	58/100 (58.0 [47.7-67.8])	<0.001
Neonatal, early onset	43/236 (12.8 [9.4-16.8])	9/100 (9.0 [4.2-16.4])	0.033
Neonatal, late onset	1/236 (0.3 [0-1.6])	2/100 (2.0 [0.2-7.0])	0.212

Note:
Numeric values are shown as proportions with exact binomial 95% CI

3.4.7 Focus of infection

Figure 3.11 shows the suspected focus of infection for all BSI episodes, stratified by age group. Despite the fact that these categories were decided based on clinically suspected focus, without the need for confirmatory diagnostics, 'non-focal infection' and 'unknown' still featured highly, at 31.8% (106/336) and 13.7% (46/336) respectively. Suspected urinary tract infections (either catheter or non-catheter associated) were important in adults, representing 15.8 % (53/336) of BSI. In children, non-focal, post-operative infection was important at 12.8% (43/336). 'Gut translocation and non-focal' was a category used in a small proportion of adults who had advanced HIV (4/158 [2.5%]) and in some children admitted with malnutrition or diarrhoea (17/178 [9.5%]).

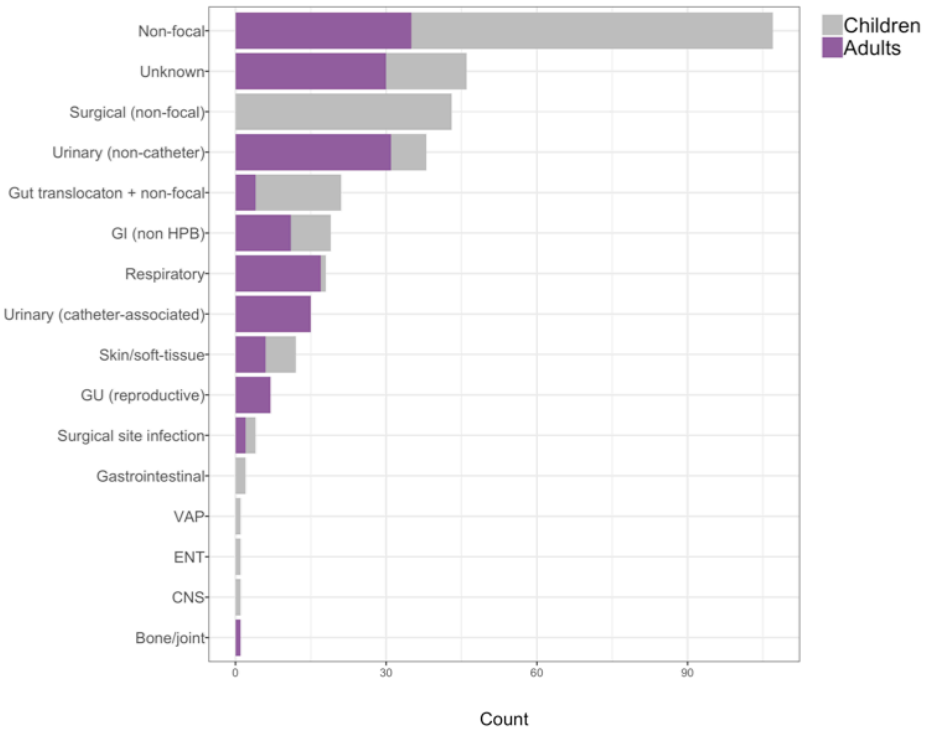


Figure 3.11: Clinically suspected focus of infection, stratified by age group.
Adults = participants aged 16 years and over, children = participants younger than 16

3.4.8 Healthcare and pre-hospital exposures

Almost two-thirds of participants had sought healthcare for their current illness, prior to attending QECH (60.1% [180/297]) (Table 3.8), which is unsurprising since referral is a requirement for adult patients. Of these, 37.8% (68/180) of participants had been admitted to another healthcare facility immediately prior to QECH and this was associated with 3GC-R on univariate analysis (OR 2.93 95% CI 1.49-6.34).

Exposure to antimicrobials for the current illness, prior to QECH presentation, was common at 29.7% (92/310) overall, with no association found between prior antimicrobial use and 3GC-R status. Although the numbers were small (Table 3.7), leading to an imprecise estimate, there was evidence of an association between prior 3GC usage for the current illness, and 3GC-R (OR 10.19 05% CI 2.09-183.55). Doxycycline usage on the other hand, showed negative correlation with 3GC-R (0.08 95% CI 0.004-0.51), again with wide confidence intervals.

Operations and urinary catheters in the three-months prior to BSI onset were common healthcare exposures and were significantly associated with 3GC-R BSI on univariate analysis (operations, OR 5.00 [95% CI 2.34-12.4] and catheters OR 2.14 [95% CI 1.26-3.77]).

Foreign travel, which is an important risk factor for acquisition of 3GC-R in other settings, was a rare exposure in this cohort, with only 10/336 participants known to have left the country within the prior 12 months.

Table 3.8: Healthcare exposures and travel, stratified by 3GC susceptibility status.

Variable	3GC-R	3GC-S	p
Pre-hospital health seeking			
Sought care prior to QECH attendance	133/208 (63.9 [57.0-70.5])	47/89 (52.8) (52.8 [41.9-61.2])	0.899
Health centre	39/133 (29.3 [21.8-37.8])	31/47 (65.9 [50.7-79.1])	-
Other government hospital	72/133 (54.1 [45.3-62.8])	8/47 (17.0 [7.6-30.9])	-
Private facility	5/133 (3.8 [1.2-8.6])	6/47 (12.8 [4.8-25.7])	-
QECH outpatient clinic	1/133 (0.8 [0-4.1])	0/47 (0 [0-0.1])	-
Non-governmental charity facility	16/133 (12.0 [7.0-18.8])	2/47 (4.3 [0.5-14.5])	-
Did not seek healthcare prior	75/208 (36.1 [29.5-43.0])	42/89 (47.2 [36.5-58.1])	-
Unknown	28/236 (11.9 [8.0-16.7])	11/100 (11.0 [5.6-18.8])	0.999
Admission to health-facility prior	**58/236 (24.6 [19.2-30.6])**	**10/100 (10.0 [4.9-17.6])**	**0.003**
Health centre	4/58 (6.9 [1.9-16.7])	2/10 (20.0 [2.5-55.6])	-
Other government hospital	41/58 (70.7 [57.3-81.9])	6/10 (60.0 [26.2-87.8])	-
Private clinic	1/58 (1.7 [0-9.2])	1/10 (10.0 [0.2-44.5])	-
Non-governmental charity facility	12/58 (20.7 [11.1-33.3])	1/10 (10.0 [0.2-44.5])	-
No admission	178/236 (72.1 [65.5-78.1])	90/100 (88.8 [80.3-94.8])	-
Other health seeking			
Healthcare attendance in prior 3m	87/153 (56.9 [48.6-64.8])	50/76 (65.8 [54.0-76.3])	0.202
No attendances	66/153 (43.1 [35.2-51.4])	26/76 (34.2 [23.7-46.0])	-
Unknown	83/236 (35.1 [29.1-41.6])	24/100 (24.0 [16.0-33.6])	0.405

Table 3.8: Healthcare exposures and travel stratified by 3GC susceptibility status (continued).

Variable	3GC-R	3GC-S	p
Pre-hospital antimicrobials			
At least one dose	64/215 (29.8 [23.7-36.4])	28/95 (29.5 [20.6-38.7])	0.999
Unknown	21/235 (8.9 [5.6-36.4])	5/100 (5.0 [1.6-11.3])	0.864
Amoxicillin†	7/215 (3.3 [1.3-6.6])	5/95 (5.2 [1.7-11.9])	0.502
Benzylpenicillin	18/215 (8.4 [5.0-12.9])	8/95 (8.4 [3.7-15.9])	0.999
Ceftriaxone/Cefotaxime	**23/215 (10.7 [6.9-15.6])**	**1/95 (1.1 [0.02-5.7])**	**0.024**
Ciprofloxacin	10/215 (4.7 [2.3-8.4])	2/95 (2.1 [0.3-7.4])	0.333
Co-amoxiclav	0/215 (0[0-1.7])	1/95 (1.1 [0.02-5.7])	0.304
Cotrimoxazole	2/215 (0.9 [0.1-3.3])	2/95 (2.1 [0.3-7.4])	0.583
Doxycycline	**1/215 (0.4 [0.01-2.6])**	**4/95 (4.2 [1.2-10.4])**	**0.029**
Erythromycin	2/215 (0.9 [0.1-3.3])	1/95 (1.1 [0.02-5.7])	0.999
Gentamicin	15/215 (7.0 [4.0-11.2])	5/95 (5.2 [1.7-11.9])	0.784
Other (includes unknown class)	8/215 (3.7 [16.2-7.2])	2/95 (2.1 [0.3-7.4])	0.999
Days prior that antimicrobials started	6.5 (2.0-19.0)	6.0 (2.0-12.8)	0.651
Other pre-hospital exposures			
Other antibiotic in prior 3 months	75/198 (37.9 [32.1-45.0])	41/84 (48.8 [37.7-60.0])	0.203
Procedures in 3m prior to BSI onset	38/226 (16.8 [12.2-23.3)	16/100 (16.0 [9.4-24.7])	0.999
Operations	**66/232 (28.4 [22.7-34.7])**	**7/95 (7.4 [3.0-14.6])**	**<0.001**
Unknown	4/236 (1.7 [0.4-4.3])	5/100 (5.0 [1.6-11.3])	0.865
Urinary catheter	**88/226 (38.9 [32.5-45.6])**	**22/96 (22.9 [15.0-32.6])**	**<0.001**
Unknown	10/236 (4.2 [2.1-7.7])	4/100 (8.0 [1.1-10.0])	0.999
Travel outside Malawi in prior 12months	5/156 (3.2 [3.0-16.4])	5/76 (8.9 [3.0-16.4])	0.303

Note: † Sum of numerators for antimicrobial classes is greater than 64 because some participants received more than one antibiotic class. Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR. Variables for which p<0.2 on univariable analysis, are shown in bold. QECH = Queen Elizabth Central Hospital

3.4.9 Household level variables and wealth indicators

Results of univariable logistic regressions using household level variables are shown in Table 3.9. These data are only available for participants who were alive at the time of recruitment (68.2% [229/336]), since they required direct questioning of participants and their families. Households had a median of 5 people living in 3 rooms. Over one-third of households kept at least one type of animal (37.5% [82/229]) and there was evidence of an association between keeping poultry and 3GC-R (OR 2.28 95% CI 1.19-4.59). Keeping livestock was non-significantly associated (OR 1.96 95% CI 0.80-5.52), potentially relating to small overall numbers (28/153 [18.3%]).

Access to flush toilets was rare (9.6% [22/229]), with most households using an ordinary pit latrine (90.0% [206/229]). There was some evidence that participants with 3GC-R BSI were less likely to be from households which shared a toilet (of any type) with other households, than from those with access to their own toilet (OR 0.55 95% CI 0.31-0.96). Just over half of households accessed drinking water from an un-piped supply (borehole/tube-well, unprotected spring/well or river) (52.0% [119/229]), with the remainder accessing water from pipes supplies inside or outside the house. There was evidence that a piped supply outside the dwelling was negatively correlated with 3GC-R (OR 0.31 95% CI 0.13-0.72) and a non-significant negative correlation between use of communal tap/standpipe water and 3GC-R (OR 0.55 95% CI 0.29-1.03). There was no such association between having piped water into the house and 3GC-R, though overall numbers were small (12.6% [22/229]).
By contrast, accessing water from a borehole was potentially associated with 3GC-R (OR 2.36 95% CI 1.32-4.31).

Households were poor, with low overall median wealth score of 2.03 (IQR 1.67-2.43), which is consistent with other studies from this setting [171]. There were, however, a lot of missing data for the covariates used in the wealth model, and it was only possible to construct scores for 47.0% (158/336) of households. As so much data were missing, I examined another potential indicator of poverty based on food security. Overall, 43.7% (100/229) of families reported problems getting the food they needed for their household over the preceding month, and there was no significant association between this variable and 3GC susceptibility status (p=0.398).

Table 3.9: Household level variables, stratified by 3GC susceptibility status.

	3GC-R	3GC-S	p
No. in household	5.0 (3.0-6.0)	5.0 (4.0-6.0)	0.860
No. rooms in house	3 (2-4)	3 (2-4)	0.889
Animals at home			
Dogs	6/153 (3.9 [1.5-8.3])	6/76 (7.9 [3.0-16.4])	0.213
Poultry	**52/153 (33.9 [26.5-42.1])**	**14/76 (18.4 [10.5-29.0])**	**0.016**
Livestock	**22/153 (14.4 [9.2-21.0])**	**6/76 (7.9 [3.0-16.4])**	**0.164**
Other animals	2/153 (1.3 [0.2-5.0])	2/76 (2.6 [0.3-9.1])	0.480
No animal	89/153 (58.1 [49.9-66.1])	54/76 (71.0 [59.5-80.9])	-
Unknown	83/236 (35.1 [29.1-41.6])	24/100 (24.0 [16.0-33.6])	0.059
WASH			
Toilet type			
Flushing toilet	15/153 (9.8 [5.6-15.7])	7/76 (9.2 [3.8-15.7])	0.886
VIP	0/153 (0 [0-2.4])	0/76 (0 [0-0.05])	0.999
Ordinary pit latrine	138/153 (90.2 [84.3-94.4])	68/76 (89.5 [80.3-95.3])	0.864
Composting	0/153 (0 [0-2.4])	1/76 (0.01 [0-0.07])	0.986
Unknown	83/236 (35.1 [29.1-41.6])	24/100 (24.0 [16.0-33.6])	0.055
Shared toilet†	**64/153 (41.8 [33.9-50.1])**	**43/76 (56.6 [44.7-67.9])**	**0.069**
Unknown	83/236 (35.1 [29.1-41.6])	24/100 (24.0 [16.0-33.6])	0.055
Households shared with†	3 (2-3)	3 (2-4)	0.538
Water for drinking			
Piped into dwelling	20/153 (13.0 [8.2-19.5])	9/76 (11.8 [5.6-21.3])	1
Piped outside dwelling	**11/153 (7.2 [3.6-12.5])**	**15/76 (19.7 [11.5-30.4])**	**0.007**
Communal tap/standpipe	**31/153 (20.3 [14.2-27.5])**	**24/76 (31.6 [21.4-43.3])**	**0.043**
Borehole/tube well	**75/153 (49.0 [40.9-57.2])**	**22/76 (28.9 [19.1-40.5])**	**0.003**
Unprotected well/spring	15/153 (9.8 [5.6-15.7])	6/76 (7.9 [3.0-16.4])	0.618
River	1/153 (0.6 [0.02-3.6])	0/76 (0 [0-0.05])	0.987
Treated drinking water	11/153 (7.2 [3.6-12.5])	6/76 (7.9 [3.0-16.4])	0.995
Unknown	83/236	24/100	0.055
Drinking water storage			
None stored in house	1/153 (0.6 [0.02-3.6])	1/76 (1.3 [0.03-0.07])	1
Plastic container	146/153 (96.7 [92.5-98.9])	70/76 (89.5 [80.3-95.3])	0.312
Metal container	0	1/76 (1.3 [0.03-0.07])	1
Clay pot	6/153 (3.9 [1.5-8.3])	4/76 (5.2 [1.5-12.9])	0.734
Unknown	83/236 (35.1 [29.1-41.6])	24/100 (24.0 [16.0-33.6])	0.055
Household wealth			
Wealth indicator score††	1.95 (1.64-2.36)	2.15 [1.81-2.53])	0.441
Difficulty feeding household*	70/153 (45.8 [37.6-54.0[)	30/76 (39.5 [28.4-51.4])	0.398

Note: † Refers to households sharing a communal toilet facility and number of households shared with †† Wealth scores missing for 158 households. * Data missing for 107 households. Numeric variables are shown as proportions with exact binomial 95%CI or as medians with IQR. Variables for which p<0.2 on univariable analysis are shown in bold. WASH = Water, sanitation and hygiene; VIP = ventilator improved pit-latrine

3.4.10 Young infants and children

A number of covariates specific to young children were explored and these are shown in Table 3.10. Numbers in these age groups were limited, with 100 young infants (aged 0-3months), including 75 neonates (aged 28 days or under). In particular, there were only 17 young infants who had 3GC-S BSI, making investigation of associations with 3GC susceptibility status difficult due to lack of heterogeneity.

Children aged 6 months to 5 years (in whom z-scores are validated)[167] were underweight, with a mean adjusted weight-for-height z-score (WHZ) of -0.99 (standard deviation [SD] 1.88) in the 3GC-R group and -0.10 (SD 1.56) in the 3GC-S group. There was no significant association between WHZ and 3GC-R or between MUAC and 3GC-R.

Children were of low birth weight in both groups (3GC-R median=1.9kg and 3GC-S median=2.3kg) with no association between birth weight and 3GC-R status (p=0.761). In total, half of young infants were born pre-term (defined as born alive before 37 weeks of completed pregnancy), but this did not appear to be a severely preterm cohort, with an overall median gestational age (GA) of 36 weeks (IQR 32-41). There was no significant association between GA and 3GC-R status (p=0.948).

Overall, 33.0% (30/91) of young infants were born at QECH with 60.4% (55/91) born in other hospitals or health centres, and 6.6% (6/91) born at home. These data were missing for 9 children. There were more young infants born in external health centres/hospitals in the 3GC-R than the 3GC-S group, but this association did not reach significance (p=0.161).

Most young infants were born by spontaneous vaginal delivery (SVD) (78.0% [78/100]), but no associations were detected between delivery mode and 3GC susceptibility status. A higher proportion of infants had received nasogastric feeding in the 3GC-R group than the 3GC-S group, but again this did not reach significance (p=0.065).

Table 3.10: Covariates unique to young children, stratified by 3GC-R susceptibility status.

Variable	3CC-R	3GC-S	p
Anthropometry†			
MUAC	12.8 (11.4-14.1)	14.7 (12.3-16.0)	0.172
Adjusted WHZ, Mean (SD)	-0.99 (1.88)	-0.10 (1.56)	0.438
Birth details			
Birthweight (kg)	1.9 (1.5-2.5)	2.3 (1.7-2.8)	0.761
Born preterm†	38/81 (46.9 [35.7-58.3])	8/16 (50.0 [24.7-75.3])	0.821
Gestational age (weeks)	35 (32-38)	36 (32-36)	0.948
Place of birth (infants 0-3m)			
Home	4/76 (5.3 [1.5-12.9])	2/15 (11.8 [1.5-36.4])	0.279
QECH	23/76 (30.3 [20.2-41.9])	7/15 (41.2 [18.4-67.1])	0.394
Other hospital or health	49/76 (64.5 [52.7-75.1])	6/15 (35.3 [14.2-76.7])	0.161
centre	7/83 (8.4 [3.5-16.6])	2/17 (11.1 [1.4-34.7])	0.660
Unknown			
Mode of delivery	66/83 (79.5 [69.2-87.6])	12/17 (70.5 [44.0-89.7])	0.752
SVD	14/83 (16.9 [9.5-26.7])	5/17 (29.4 [10.3-56.0])	0.322
Caesarian	3/83 (3.6[0.1-10.2])	0/17 (0 [0-19.5])	0.999
Assisted			
Feeding	18/83 (21.7[13.3-32.21])	5/17 (29.4 [10.3-56.0])	0.532
Breast	3/83 (3.6[0.8-10.2])	1/17 (5.9 [0.01-28.7])	0.531
Bottle	23/83 (27.7[18.4-38.6])	1/17 (5.9 [0.01-28.7])	0.065
Nasogastric	39/83 (47.0[35.9-58.2])	10/17 (58.8 [32.9-81.6])	0.431
Mixed			

Note:

† Anthropometry is for children aged 6months to 5 years. The remaining variables are for children ≤ 3 months

Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR.

MUAC = Mid-upper arm circumference, WHZ = weight-for-height Z.score, SD = standard deviation, QECH = Queen Elizabeth Central Hospital, SVD = Spontaneous vaginal delivery

3.4.11 Predictors of 3GC-R BSI

3.4.11.1 Summary of univariable associations with 3GC-R BSI

Variables for which p<0.2 on univariable analysis, which were taken forward for analysis in the multivariable model, are shown in Table 3.11, stratified by patient and household level factors, and healthcare exposures. Variables for which the 95% confidence intervals of odds ratios do not cross 1 are shown in bold.

Table 3.11 Univariable associations with 3GC-R in bloodstream infection.

Variable	Univariable models OR (95% CI)
Patient factors	
Age	**0.98 (0.97-0.99)**
HIV infected (>18months old)	**0.33 (0.18-0.58)**
HIV exposed (<18months old)	2.50 (0.65-16.4)
HIV infected on ART	**0.25 (0.15-0.44)**
HIV infected on CPT	**0.34 (0.20-0.58)**
Temperature on admission (°C)	**0.83 (0.70-0.97)**
Unable to stand on admission	**2.21 (1.24-3.92)**
Unable to stand on recruitment	**2.65 (1.48-4.77)**
Healthcare exposures	
Prior admission to health centre	**2.93 (1.49-6.34)**
Ceftriaxone before admission	**10.19 (2.09-183.55)**
Doxycycline before admission	**0.08 (0.004-0.51)**
Operations in the prior 1 month	**5.00 (2.34-12.4)**
Catheter in the prior 1 month	**2.14 (1.26-3.77)**
Household factors	
Household keeps poultry	**2.28 (1.19-4.59)**
Household keeps livestock	1.96 (0.80-5.52)
Household shares toilet	**0.55 (0.31-0.96)**
Drinking water piped outside	**0.31 (0.13-0.72)**
Drinking water from communal tap or standpipe	0.55 (0.29-1.03)
Drinking water from borehole or tube well	**2.36 (1.32-4.31)**

Note: OR were calculated from univariable logistic regression models with each variable alone considered as a predictor. OR for which the 95% CI do not cross 1 are shown in bold. ART = Antiretroviral therapy, CPT = Cotrimoxazole preventive therapy

The amount of missing data is a problem for these univariate estimates, which is compounded when attempting to construct a multivariable model. Proportions of missing data for each variable are shown in the appendix to this chapter (Appendix Figure 3.1). Most missing data points occurred as a result of recruiting patients who had died by the time of recruitment ([107/336], 31.8%), but excluding these participants from a risk-factor model would exclude the sickest patients, introducing considerable bias to the estimates. An attempt at constructing a model using complete case analysis, removed 129/336 data rows and was therefore subject to considerable bias. Excluding variables with missing data would leave only seven variables left in the model (Appendix Figure 3.1).

3.4.11.2 Multiple imputation of missing data

To correct for missing data bias and increase the size of the available dataset, I imputed missing data using multiple imputation of chained equations in the 'mice' package in R[172]. This fully conditional specification approach, imputes missing values based on observed values for a given individual and the relationships between observations in the data for other participants. It produces a "complete" data set, with missing covariate values replaced by imputed values. By imputing multiple times, multiple predictions are created for each missing value, helping to account for uncertainty in the estimates. Default settings in 'mice' were used, to produce 100 imputed datasets. Using the mice algorithm, logistic regression models are then fit using all imputed datasets and pooled posterior parameter estimates calculated[173].

3.4.11.3 Inspection of imputed data

The plausibility of imputed data was assessed by comparing the density distributions of observed and imputed data (Figure 3.12). The density distributions for each imputed variable with observed data are similar and overlapping[174]. Using 100 iterations of the mice algorithm, plausible datasets were therefore created.

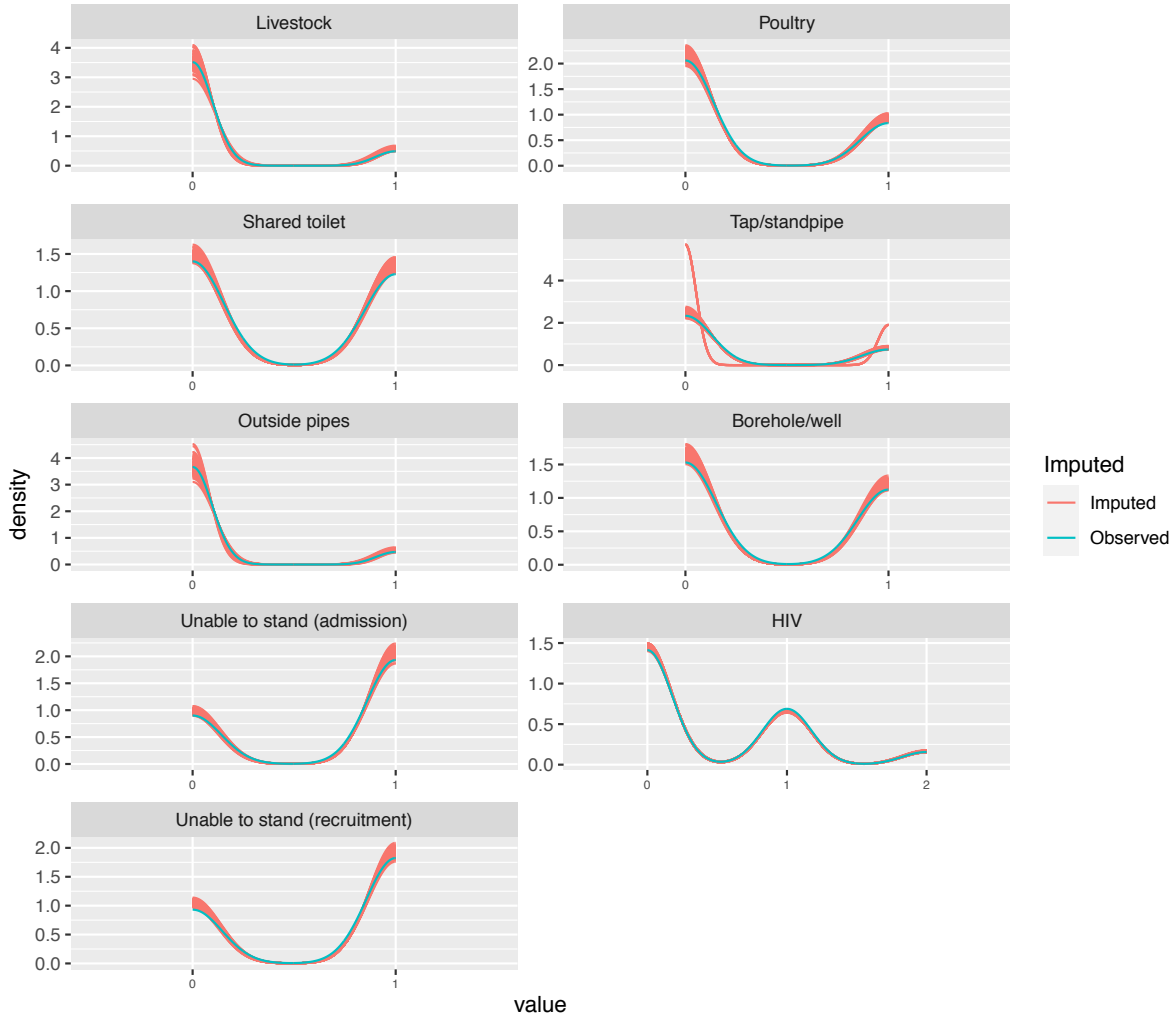


Figure 3.12: Density distributions of observed and 100 imputed datasets, shown for all discrete variables in which data were imputed. The density distributions of imputed data are similar and overlapping, suggesting plausibility of imputations. Note HIV variable coded as factors: 1 HIV infected, 2 HIV exposed, 0 HIV negative.

3.4.11.4 Variable selection

The imputed data described above was then explored as follows. A potential concern was model over-fitting resulting from the large number of parameters selected from univariable analysis. An over-fitted model would describe random error in the data, rather than the true relationship between variables[175]. I therefore explored two methods to reduce the dimensionality of the dataset by collapsing predictor variables into a smaller number of variables, to select the best subset of predictors:

1. Inspection for collinearity and;
2. Stepwise backward elimination.

The methods and results of each are presented below.

3.4.11.4.1 Inspection for collinearity and a priori model

Bivariate analysis was conducted by construction of a non-parametric (Spearman) correlation matrix (Figure 3.13). Pairs of variables which were highly correlated (correlation coefficient >0.9 and which could feasibly be duplicating the effect on outcome variance were examined)[175]. A variable in each highly correlated pair was removed based on a priori knowledge about their relationship and the amount of observed data available for each variable in the pair. For example, having HIV and taking ART were unsurprisingly highly correlated (correlation coefficient 0.96). ART was removed as a predictor variable because this is entirely driven by being HIV infected and there was no evidence that the negative association of HIV on 3GC-R was mediated by ART, as both associations persisted on a multivariable model using both HIV and ART as predictors.

3.4.11.4.2 Stepwise backward elimination

This stepwise selection method iteratively removes predictor variables, in order to find the subset of variables in the data which result in the best performing model, using Akaike information criterion (AIC), as a measure of model fit. A backward elimination algorithm was built using custom code in R, taking multiply imputed data from the mice output (see Chapter Appendix). This algorithm calculates a pooled AIC by taking the arithmetic mean of AICs from each realisation of the imputed dataset, and calculates a pooled AIC for a model fit with each explanatory variable removed in turn. If a drop in AIC occurs between iterative models, the variable that gives the greatest drop is removed and a new AIC calculated for that model. If no drop in AIC occurs between steps, then the algorithm stops to leave the most parsimonious model.

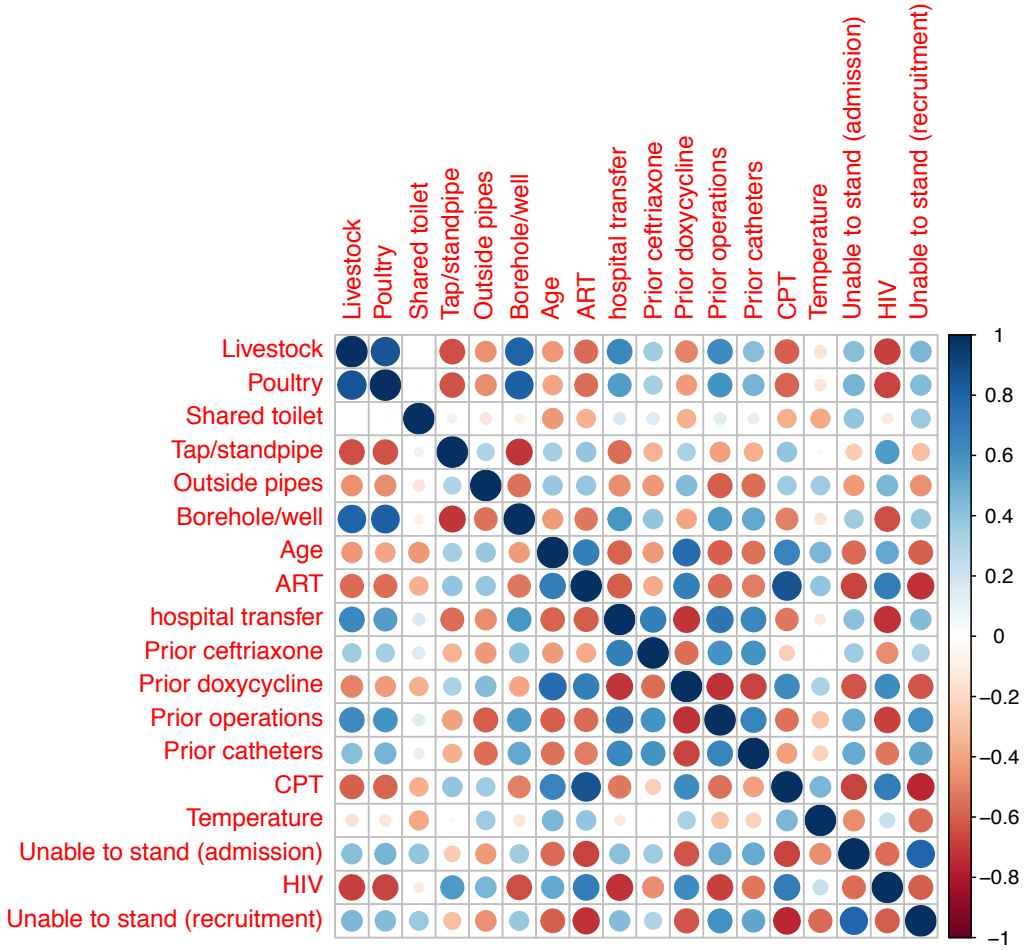


Figure 3.13 Visual representation of non-parametric correlation matrix for predictor variables. Colour intensity and size of circle are proportional to the correlation coefficients. Blue shades = positive correlation, red shades = negative correlation. Legend shows relationship between colour and correlation coefficients.

3.4.11.5 Final model predictions

Table 3.12 shows the final selected predictor variables for both of the approaches described above, with adjusted odds ratios (aOR) calculated from the two final multivariable logistic regression models. The final a priori model contained two predictor variables which were not in the stepwise model (water from borehole/well and unable to stand on admission). The stepwise model contained two variables which were not in the a priori model (ceftriaxone before admission and prior admission to

health centre). The AIC in the stepwise model was lower, suggesting a better overall fit. Reassuringly, the final multivariable associations with 3GC-R were the same in both models. Prior operations were associated with 3GC-R and shared toilets, HIV infection and older age were negatively associated. The large effect size for the association of prior ceftriaxone and HIV exposure with 3GC-R from the stepwise model are notable, though the confidence ratios for the aOR are wide and cross 1 in the case of ceftriaxone.

Table 3.12: Multivariable associations with 3GC-R bloodstream infection, shown for both final models.

Variable	a priori model OR (95% CI)	p-value	Stepwise model OR (95% CI)	p-value
Patient factors				
Temperature	0.98 (0.95-1.01)	0.139	0.83 (0.68-1.00)	0.053
HIV	0.84 (0.75-0.96)	0.046	0.44 (0.24-0.81)	0.008
HIV exposed	1.12 (0.93-1.36)	0.179	3.48 (0.70-17.27)	0.127
Age	**0.99 (0.97-0.99)**	**0.018**	**0.99 (0.98-0.99)**	**0.029**
Healthcare exposures				
Prior operations	**1.15 (1.02-1.29)**	**0.046**	**2.71 (1.13-6.65)**	**0.025**
Unable to stand (admission)	1.01 (0.89 – 1.15)	0.589	-	
Prior ceftriaxone	-		6.34 (0.73-54.4)	0.092
Hospital transfer	-		1.80 (0.78-4.18)	0.170
Household factors				
Shared toilet	**0.89 (0.79-0.99)**	**0.029**	**0.46 (0.23-0.90)**	**0.023**
Outside pipes	0.85 (0.71-1.03)	0.159	0.51 (0.19-1.35)	0.172
Borehole/well	0.79 (0.31-1.93)	0.811	-	-
Poultry	1.08 (0.95-1.21)	0.235	1.79 (0.82-3.89)	0.142

Note:
OR were calculated from multivariable logistic regression models.
OR for which the 95% CI do not cross 1 are shown in bold.

3.5 Discussion

3.5.1 Characteristics of patients with bloodstream Enterobacterales and *Acinetobacter*.

In this chapter I have presented a detailed clinical and microbiological description of patients who have bloodstream infection with Enterobacterales in Blantyre, Malawi and attempted to understand the risk-factors for 3GC-R in these infections.

This study corroborates recent findings from Blantyre, by demonstrating concerningly high levels of 3GC-R amongst key bloodstream Enterobacterales in patients presenting to QECH[3]. In this cohort, over 70% of Gram-negative BSI were resistant to ceftriaxone and almost half of all were resistant to all readily available antimicrobials. Whilst long-term surveillance of BSI at QECH has documented the emergence and spread of 3GC-R amongst Enterobacterales, occurring since the roll-out of ceftriaxone in 2005, this study is the first to link these diagnostic laboratory isolates to prospectively collected clinical data, and therefore allow a deeper understanding of these patients.

In European cohorts, Gram-negative sepsis tends to affect older age groups[160, 176], but in keeping with other sepsis cohorts from Blantyre[177, 178], the patients in the current study were young, even when considering the adult cohort separately. Malawi national treatment guidelines state that ART should be started as soon as possible for all adults and children with confirmed HIV infection, using a dolutegravir-based regimen (with tenofovir/lamivudine backbone). CPT should be

started for all HIV infected adults and for HIV exposed or infected children from the age of 6 weeks[152]. CPT should be continued lifelong in adults and in HIV exposed children when they are confirmed negative and breastfeeding has stopped[152]. In this cohort, although reported ART coverage was high, and although difficult to interpret in the context of acute illness, CD4 counts were low, perhaps reflecting that bloodstream infection is associated with ART failure. Data on ART failure in Malawi are limited, but virological failure appears to be common amongst hospitalised inpatients in Malawi (32% of patients taking ART for > 6 months in one cohort study).

BSI is a common cause of morbidity and mortality in sSA[179], but there are no other large-scale studies of Gram-negative sepsis from this setting, to which these results can be compared.

Limited by a lack of robust diagnostics, the underlying focus of bloodstream infection was difficult to elucidate in these patients. In other settings, urinary tract and hepatobiliary infection predominate[160], but lack of timely urine culture and reliable ultrasound mean that these diagnoses may be underestimated in this cohort. Largely based on clinical syndrome, however, UTIs (both catheter associated and community acquired) were important and reliable front-door urinary culture would likely reveal many more of these infections. Our urine samples, were collected a median of 8 days post-admission, limiting the reliability of the results (only 4 samples were culture positive for the same organism as the BSI). Even in studies from high-income settings, the source of BSI is

unknown in about 25% of cases[27], and there is some evidence that this is associated with poor outcomes[180, 181].

Based on vital signs and blood parameters, patients in this cohort appeared have less severe sepsis than might be expected. These results must, however, be treated with caution due to the amount of missing data and the time period from admission to blood sampling. Participants who had died before being identified for recruitment could not have recruitment physiology or blood tests recorded. Likewise, because participants were not recruited at the front door, admission physiology is that which is recorded as part of routine hospital care – therefore prone to errors and gaps. Blood tests and vital signs on recruitment were taken a median of 8 days after participants were admitted, allowing for improvement or normalisation of parameters with treatment.

Prior to this study, it was assumed that most bloodstream infections from patients at QECH were community acquired[3], because of the general policy that patients have one blood culture on presentation and no follow-up blood cultures. By characterising each BSI episode in relation to pre-hospital healthcare exposures and the timing of blood culture in relation to admission, I show that over 60% if Gram-negative BSI occur in relation to hospital or other healthcare. This is clearly important when empirical treatment regimens for these infections.

3.5.2 Risk-factors for 3GC-R in bloodstream infections: which patients may need a different empirical antibiotic?

Ceftriaxone was rolled out in Malawi in 2005, partly in response to a high burden of invasive bacterial infection caused by multi-drug resistant Salmonellae, and is now relied upon for the treatment of sepsis in adults and children at QECH. In the chapter that follows, I will review the specific antibiotic regimens given to patients in this cohort, but access to alternative antibiotics that retain activity against 3GC-R is severely restricted, and knowing which patients at QECH are likely to be at risk from 3GC-R infections is therefore critical.

As in other settings, healthcare exposures were potentially important in these patients[27, 181, 182]. More healthcare associated BSI were 3GC-R than 3GC-S and prior admission to a healthcare facility was associated with 3GC-R BSI on univariate analysis, as were surgery and urinary catheters in the one-month prior to BSI. Ceftriaxone prior to admission showed a strong association with 3CG-R (OR 10.19 95% CI 2.09-183.55), the wide confidence intervals likely due to the small overall number of subjects on ceftriaxone. Although only the prior operations association persisted on multivariable analysis, clinicians should review their patient's pre-hospital antibiotic regimens and exposures prior to re-starting empirical ceftriaxone at QECH. A potentially protective effect of doxycycline is difficult to interpret and may simply be an artefact of the very small sample size – 5 participants in the 3GC-S group and 1 participant in the 3GC-R group. It is possible that this simply reflects reduced use of 3GC in these patients and that use of alternative broad-

spectrum antibiotics is protective against 3GC-R, but again the small effect size is difficult to interpret.

A sustained outbreak of post-operative BSI with 3GC-R K. pneumoniae in young infants operated at the MJC paediatric surgical hospital, accounts for the significant association of surgery with 3CG-R BSI, which persisted on multivariable analysis. Indeed, antibiotic regimens for these patients have already been reviewed and where available, amikacin has been incorporated into empirical post-operative sepsis regimens.

The potential impact of WASH factors on 3GC-R are interesting. Accessing the main source of drinking water from a piped supply was potentially protective for 3GC-R on univariable analysis. Water from a communal tap or standpipe, another type of piped supply, also showed negative correlation, though with confidence intervals crossing 1 (OR 0.55 95% CI 0.29-1.03). Both of these water supplies originate from Blantyre water board provided water and are therefore chlorinated. In contrast, using borehole/tube-well water as the main drinking water source, was positively associated with 3GC-R. These latter water supplies are frequently poorly constructed and maintained, and therefore more likely to be contaminated than the piped supplies. The protective effect of shared toilets is harder to interpret, as shared sanitation is typically associated with adverse health outcomes when compared to individual household latrines[183]. However, some studies from Africa find shared sanitation facilities to provide a protective affect against diarrhoea, particularly in countries where private sanitation is less available[184] and this association clearly warrants further investigation in Blantyre.

For the association of these household level covariates with 3GC-R BSI to be plausible, the assumption is that they are first driving gut mucosal colonisation, which is then a risk factor for subsequent invasive infection. Although I do not have pre-admission stool samples to show this association, it is a well described phenomenon in other contexts[27], and I will explore risk-factors for faecal 3GC-R carriage in Chapter 6.

Perhaps surprising, was the negative association of HIV with 3GC-R BSI in participants over 18 months of age, which persisted on multivariable analysis. Immune suppression from diabetes is a risk factor for 3GC-R[185], possibly driven by more frequent hospital attendance and antibiotic use, but there are surprisingly few data exploring the effects of HIV on 3GCR-E carriage[29] or infection[16]. Studies exploring risk factors for 3GC-R in sSA are in children, in whom HIV prevalence is generally too low to determine an effect[92, 103]. However, one large retrospective study from South Africa, found that other chronic disease, but not HIV were risk factors for 3GC-R *E. coli* BSI[186]. It may be that in this cohort, there was not enough heterogeneity in HIV status to determine a true effect, but this association clearly warrants further investigation in this and other settings with a generalised HIV epidemic.

In children under 18 months of age, in whom HIV infection cannot be confirmed on antibody testing (HIV PCR was not done as part of the study), HIV exposure was non-significantly associated with 3CG-R (OR 2.50 95% CI 0.65-16.4)). The small number of HIV exposed infants in the cohort makes this an imprecise estimate, and lack of confirmatory HIV

testing makes this result more difficult to interpret. In Malawi, around 5-8% of these HIV exposed infants will test positive by 24 months of age[187], but there is some suggestion that even HIV exposed, uninfected children are immune compromised with higher morbidity from infections[188]. Perhaps the relative immune compromise of being very young (most children under 18 months old were neonates) and HIV exposed has a combined immune-modulatory effect that is more important than HIV alone.

Certainly, risk factors for 3GC-R in young infants and neonates are important and warrant further investigation, but this study was underpowered to explore these associations. Most young infants were born outside of QECH, and although one-third of these 'outborn' infants were referred to QECH for surgery and developed BSI in relation to the MJC outbreak, the remainder appear to have acquired their infections in health settings outside of QECH. This finding has potential implications for the empirical management of neonatal sepsis in a setting where, overall, 80% are inborn (K. Kawaza, personal communication).

3.5.3 Limitations

This study and analysis have a number of limitations. The study was not powered to detect clinically significant risk-factors for 3GC-R in BSI and this accounts for much of the imprecision and uncertainty in the estimates presented. This was further compounded by the amount of missing data on exposures (due to the high mortality prior to blood culture results). Multiple imputation allows for pooling of information across variable when data are missing, but may lead to models which are not as accurate as ones

in which data were complete. I did not recruit a non-bacteraemic control group, which might have increased the power of the study to detect significant risk factors. Whilst future studies could proceed with larger sample sizes and additional control groups, the high proportion of deaths occurring before blood culture results are available will continue to make prospective data collection and risk-factor analysis a challenge in this setting. The alternative would be a large prospective fever study, but as only approximately 10% of blood cultures are positive for any pathogen, this would require a huge investment. High early-mortality was similarly a problem for collection of accurate patient physiology and blood tests, thus limiting the interpretability of these variables in the models. In a setting such as QECH, however, where robust quality-controlled measurements are not available as part of routine clinical care, retrospective data is equally hard to interpret. Additionally, patients who were discharged home before their blood culture result was available were usually uncontactable. As a result, the study may have considerable sampling bias towards the sicker patients (amongst those who survived until blood culture results were available).

There were problems with the logistic regression models used for the risk-factor analysis, that may have led to bias in the estimates presented. In the a priori model, variables were excluded based on collinearity and expert opinion about their relationships. Removing certain variables does not necessary mean that they are not correlated with the outcome, but only that they are not required to predict the response[189]. Step-wise variable elimination can partially reduce this bias, but have been shown to underestimate the standard errors of coefficient estimates[190].

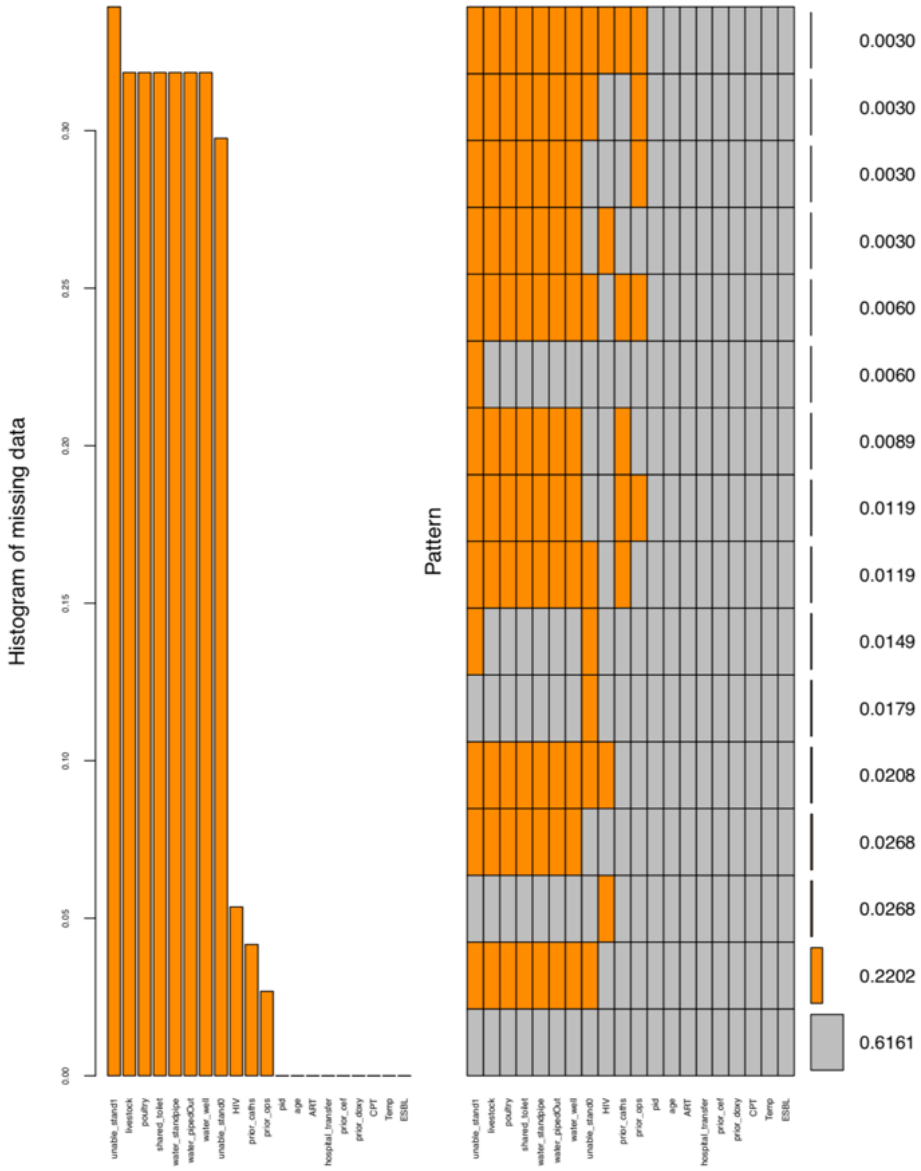
Despite these attempts at dimensionality reduction, it is still possible that the models were over-parametrised, with a large number of predictor variables for one response variable[190]. If the variables in the model are not acting independently, their inclusion in the multivariable analysis can underestimate their associations on 3GC-R[143]. For example, the relationship between catheters, operations and ceftriaxone may be causal rather than independent. To overcome this, a full causal framework expressing the conditional relationships among variables in the model would be required[191].

3.6 Conclusions and future work

Despite these limitations, I found several plausible and readily-identifiable associations between patient characteristics and 3GC-R which may be helpful for future prediction models and for clinicians deciding on empirical antibiotic regimens for their patients. The problems with these models, however, mean that further risk factor studies of 3GC-R in BSI would be valuable. Future studies could perhaps focus on a single age group, to allow a larger sample to be studied and potentially include non-bacteraemic controls. Ultimately, however, long-term high-quality AMR surveillance, implemented alongside clinical care, is what is needed to generate the unbiased data required to fully understand these infections.

Understanding risk-factors will be vital in exploring transmission and reducing 3GC-R and ultimately potentially adverse outcomes for patients. The following chapter explores these outcomes in detail, by describing the morbidity and mortality burden from 3GC-R.

3.7 Appendix



Appendix Figure 3.1: Missing data for model variables. Left-hand histogram shows proportions of missing data for each variable, in orange. Right-hand tile plot shows frequencies for different combinations of variables missing. For example, the combination of 8 variables shown in orange on the second-row from bottom are missing with a frequency of 22%

Chapter 4

Morbidity and mortality burden of third generation- cephalosporin resistant bloodstream infection in Blantyre, Malawi

4.1 Overview

In this chapter, I present the results of the primary analysis of the hospital cohort, which aimed to describe the association between 3GC-R bloodstream infection and mortality and hospital length of stay. I first present a directed acyclic graph (DAG) to visually represent the association between 3GC-R and death and use this DAG to identify a minimum adjustment set of confounders aimed at minimising the bias in the subsequent causality models. I then use logistic regression and survival analysis to estimate the association between 3GC-R on in-hospital mortality, 6-month survival and hospital length of stay.

The in-hospital case-fatality rate (CFR) for all participants in the cohort was 46.4% (156/336). Amongst participants with 3GC-R, this was 49.2% (116/236) and amongst participants with 3GC-S, this was 40.0% (40/100). There was a significant association between 3GC-R and in-hospital mortality on logistic regression (odds ratio, OR 1.85, 95% CI 1.06-3.26). In Cox proportional hazard models, 3GC-R was associated with death (hazard ratio, HR 1.51, 95%CI 1.04-2.18) and increased hospital length of stay (HR

for discharge 0.70, 95%CI 0.49-0.99). In all models, this effect was reduced when conditioning on use of effective antibiotics.

4.2 Chapter aims

The aims of this chapter are:

1. To quantify the effect of 3GC-R on in-hospital mortality and 6-month survival in patients with confirmed BSI due to Enterobacterales or *Acinetobacter* spp.

2. To quantify the effect of 3GC-R on hospital length of stay in patients with confirmed BSI due to Enterobacterales or *Acinetobacter* spp.

4.3 Methods

I first present the antibiotic treatments used for participants in the cohort, using simple proportions. I then describe an attempt to classify patients according to severity of sepsis, and explain the reasons this was problematic in this cohort.

For Aim 1, in-hospital mortality is presented as case-fatality proportions, with exact binomial 95% confidence intervals, stratified by 3GC-R status. P-values were calculated using univariable logistic regression models for mortality and 3GC-R status alone as a predictor, using the *glm* function in R.

Kaplan-Meir estimates of the survival function were then generated using the survival package in R[192]. Day 0 was day of infection, which was assumed to be the day of blood culture sampling. Observations were right-censored at the end of the study follow-up period (180-days). The log-rank test was used to test for difference in survival by 3GC-R status. Empirical survival curves were generated using the survfit function in R and are displayed aggregated and stratified by 3GC-R status.

Multivariable models which adjusted for relevant confounders were then built, to investigate the causal relationship between 3GC-R and mortality. I used a causal inference framework[193] to identify potential confounders in these models and this approach is described in more detail in Section 4.4.3 below. Briefly, a causal diagram, known as a directed acyclic graph (DAG), representing the relationships between 3GC-R and death was drawn using a causal inference framework[193]. The dagitty package in R[194] was used to automate the framework, in order to determine a minimum adjustment set of conditioning variables. The identified variables were used in subsequent logistic regression and Cox-proportional hazard models.

I first used a multivariable logistic regression model for estimation of association between 3GC-R on in-hospital mortality, using the *glm* function in R. Odds ratios are presented with 95% exact binomial confidence intervals. I then used a multivariable Cox-proportional hazards model to estimate the impact of 3GC-R on 180-day survival. The time scale for the Cox model was day of infection and observations were right-censored at

180-days. Hazard ratios for mortality from 3GC-R are presented with 95% confidence intervals.

To determine the association between 3GC-R and hospital length of stay, similar Cox models were constructed, but using day of hospital admission to day of discharge as the time-frame. Observations were right-censored if death occurred before discharge and hazard ratios for discharge presented with 95% confidence intervals.

4.4 Results

4.4.1 Antibiotic treatment

Overall, 310/336 (92.3%) of participants received at least one antibiotic during their BSI episode. The empiric antibiotic regimen (the antibiotic given prior to final culture and sensitivity results) contained an "effective antibiotic" (an antibiotic which was active against the cultured bloodstream isolate based on susceptibility testing) in 113/336 (33.6%) cases overall. An effective empirical empiric antibiotic had been given in 82/100 (82.0%) of 3GC-S BSI and in 31/226 (13.7%) of 3GC-R.

Including empiric antibiotics and regimens tailored to blood culture results, 89/100 (89.0%) participants with 3GC-S BSI received at least one dose of effective antibiotic during the course of their infection, but for patients with 3GC-R infection, this was only 90/226 (39.8%). Overall then, only 179/336 (53.3%) of participants received an effective antibiotic for their BSI. Median time to effective antibiotic was 5 days (IQR 0-7) for patients with 3GC-R BSI and 0 days (IQR 0-1) for patients with 3GC-S infection.

Figure 4.1 shows the overall antibiotics received by participants in the cohort, stratified by 3GC-R status.

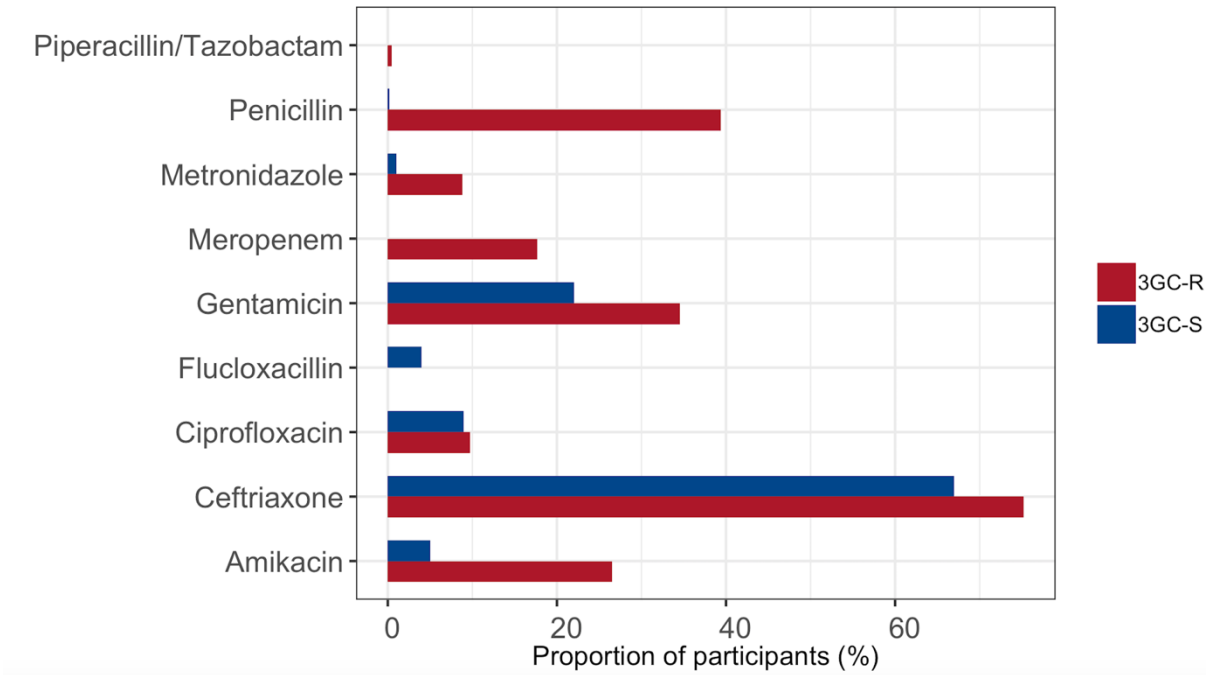


Figure 4.1: Antibiotics received by participants with BSI, stratified by 3GC-R status.

4.4.2 Sepsis severity

The application of sepsis severity scores, which are used predict poor outcomes from sepsis, is problematic in LMICs[195], and I will address this further in the discussion section of this chapter. I considered four possible sepsis classifications to define sepsis severity in this cohort: the sepsis-related organ failure assessment (SOFA)[196], the quick SOFA (qSOFA)[197], the universal vital assessment score (UVA)[198] and systemic inflammatory response syndrome (SIRS) criteria[199].

The SOFA score stratifies mortality risk in septic patients based on the degree of dysfunction in 6 organ systems and requires a large number of parameters, including blood tests such as bilirubin and the requirement for

ventilation and inotropic support, data which are not routinely available at QECH and which were not collected in this cohort. The simplified qSOFA score (any two of altered mental state, SBP <100mmHg, respiratory rate >22 breaths per minute) has been used to predict poor outcomes in high-income settings, with a score of 2 or more predictive of mortality[197]. However, few patients in this cohort had conscious level recorded on admission, with data missing for 230/336 (68.5%) of participants.

A UVA score has been derived and validated for hospitalised patients in a number of sub-Saharan African countries[198]. The score includes points for systolic blood pressure, oxygen saturation and Glasgow Coma Scale (GCS), but again, these admission data were missing in a high proportion of the cohort (systolic blood pressure missing in 175/336 [52.1%]) and the score is validated in adult populations only[198].

SIRS criteria relate to the host inflammatory response to both infectious and non-infectious stimuli. Although this terminology has largely been abandoned in newer sepsis definitions, due to its lack of predictive validity for in-hospital mortality[197], the parameters (heart rate, respiratory rate, white blood cell count and temperature) remain easy to collect in LMICs and are adaptable to all age-groups.

For adult patients over 18 years, SIRS is defined as at least two of: fever >38.0°C or hypothermia <36.0°C, tachycardia >90 beats/minute, tachypnoea >20 breaths/ minute, leucocytosis >12 x109/l or leucopoenia <4 x109/l[199]. The parameters for age-adjusted SIRS in children are shown in the appendix to this chapter (Appendix Table 1). Table 4.1

shows the percentage of patients with SIRS in this cohort, stratified by 3GC-R status. Although these parameters are simple, there are still a substantial amount of missing data. Although vital sign data at the time of recruitment is much more complete than at the time of admission (as described in Chapter 3, Table 3.6), the latter are collected a median of 8 days after admission, and therefore in most cases, do not reflect the severity of sepsis at the onset of bloodstream infection.

Table 4.1: Proportion of participants with SIRS, stratified by 3GC-R status

	3GC-R	3GC-S	Total	p
Presence of SIRS n/N (%[95%CI])	97/141 (68.8[60.5-76.3])	50/72 (69.4[57.5-79.8])	147/213 (69.0[62.3-75.1])	0.999
Missing parameters	95/236 (40.2[29.9-42.4])	28/100 (28.0[19.5-37.9])	123/336 (36.6[31.4-42.0])	0.036

Note:
SIRS = Systemic inflammatory response syndrome (defined in text, Section 4.4.2)
Three parameters for the SIRS score (Heart rate, Temperature and Respiratory rate) were recorded on admission. White cell count was measured on recruitment.

4.4.3 Causal structure of death in 3GC-R bloodstream infection

Before modelling the causal effect of 3GC-R on death in regression models, I considered the role of covariates in these models from a causal inference perspective[200], using a directed acyclic graph (DAG). DAGs are graphical tools which allow for visualisation of the key concepts in causation – namely, exposure, outcome, causation and confounding[201]. This process allows for clear definitions of which variables are acting as confounders, in order to maximally reduce bias in the causation models. Using a pre-hypothesised causal directed DAG, I hypothesise the causal relationship of 3GC-R on mortality as shown in Figure 4.2. In the DAG, nodes (such as 3GC-R age and HIV status) represent potential model

variables and arrows (or edges) represent causality, for example host factors such as age and HIV status may influence sepsis severity, 3GC-R status and presence of other pathogens.

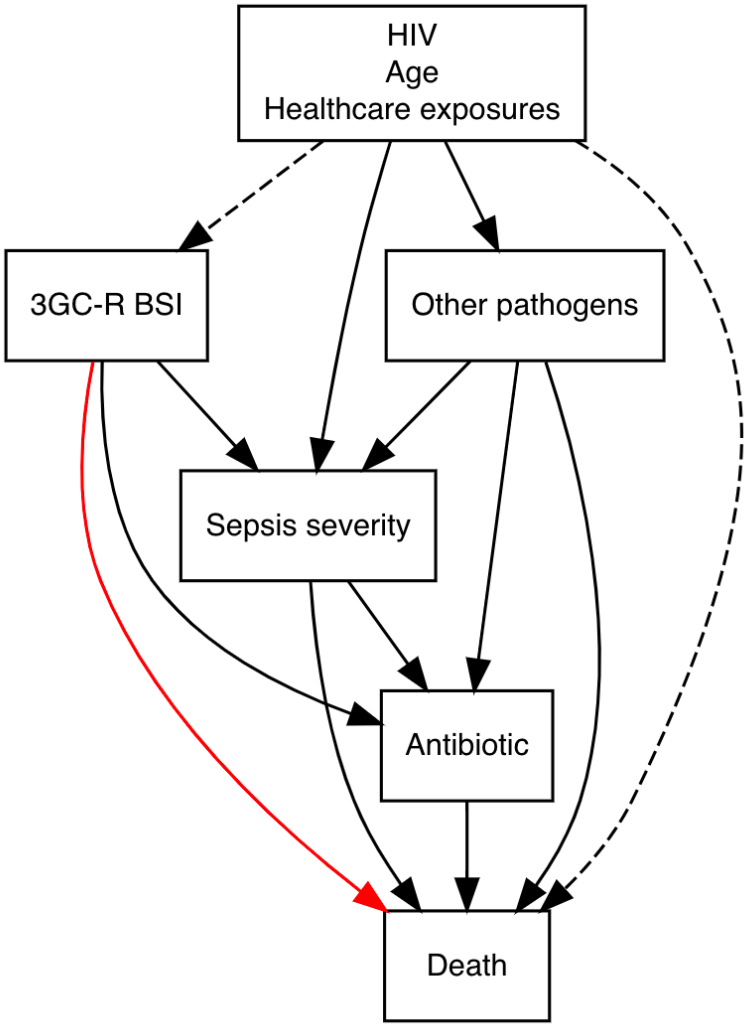


Figure 4.2: Hypothesised causal diagram for causal effect of 3GC-R BSI on death. Causal paths are shown as arrows. Host-factors such as HIV, age and prior healthcare exposure are exogenous factors affecting both the exposure (3GC-R) and outcome (Death) and are therefore potential confounded associations (biasing paths) shown as dotted arrows (or edges). Red arrow indicates direct causal effect of 3GC-R on death. Death from 3GC-R is mediated in part by sepsis which may vary in severity or be absent (Sepsis severity). Sepsis severity may influence the antibiotic given which in turn may be active against the infection, not active or not given at all. Other infectious pathogens may influence sepsis severity, treatment choice and outcome. To estimate the total effect of 3GC-R on death, only biasing pathways must be closed. To estimate the direct effect of 3GC-R on death (red arrow), all other pathways must be closed.

The aim of my analysis is to estimate the effect of 3GC-R on mortality. The purpose of the DAG is to identify a minimum adjustment set which will minimise bias when estimating this causal effect. Based on the above diagram, the minimum adjustment set for the total effect of 3GC-R on

death (closure of only biasing paths, shown as dotted edges) requires adjustment for age, HIV and healthcare exposures. To estimate the direct effect of 3GC-R on death, all other biasing and causal paths must be closed, leaving only the edge from 3GC-R to death (shown in red). To estimate the total effect of 3GC-R on death, the minimum adjustment set is therefore age, HIV and healthcare exposures. To estimate the direct effect of 3GC-R on death, would additionally require adjustment for sepsis severity, other pathogens and treatment with an effective antibiotic.

4.4.3.1 Selection of model covariates

Two models were fitted for each of the chapter aims. The first model aimed to estimate the total effect of 3GC-R on mortality/length of stay and conditioned upon age, HIV and healthcare exposures. Healthcare exposures were selected if they were associated with 3GC-R on the univariate analysis carried out in Chapter 3 and if complete data were available for the variable. These variables were therefore prior operations and prior hospitalisation.

The second model would ideally condition upon other pathogens, sepsis severity and treatment with effective antibiotics, in order to estimate the direct effect of 3GC-R on mortality/length of stay. Effective antibiotic usage data is available as described in section 4.4.1 above. This was included in the models as a binary covariate (use of effective antibiotic or not) as only 7.7% of participants received no antibiotic, meaning there would be insufficient heterogeneity to include this in the covariate. As discussed in section 4.4.2, there are too many missing data to include a sensible "sepsis severity" covariate in the models. In terms of "other

pathogens", MTB is the most likely to be present in this cohort (personal communication J. Lewis), but accurate TB diagnostics were not available at QECH at the time the study was recruiting and these data are again missing. Because of these limitations, the second model conditioned upon age, prior operations, prior hospitalisation and effective antibiotic use only.

4.4.4 Mortality burden from 3GC-R BSI

The in-hospital case-fatality rate (CFR) for all participants in the cohort was 46.4% (156/336). Amongst participants with 3GC-R, this was 49.2% (116/236) and amongst participants with 3GC-S, this was 40.0% (40/100) (p=0.125).

Table 4.2 shows in-hospital CFR stratified by age-group. These stratified estimates are limited by small sample sizes, but CFR was highest in the young infants

(< 3-months of age) at 61.0% (61/100) and no difference was seen between participants with 3GC-R and 3GC-S infection (61.4% versus 58.8%, p=0.999). The greatest difference in CFR between those with 3GC-R and 3GC-S infection, occurred in adults over 16 years (53.3% vs 39.1%, p=0.081).

Table 4.2: In-hospital case-fatality rates stratified by 3GC-R status

Age group	Case-fatality rate n/N (% [95%CI]) 3GC-R	3GC-S	p-value	Total
All	116/236 (49.2 [42.6-55.7])	40/100 (40.0 [30.3-50.3])	0.151	156/336 (46.4 [41.0-51.9])
0-3m	51/83 (61.4 [50.1-71.9])	10/17 (58.8 [32.9-81.6])	0.999	61/100 (61.0 [50.7-70.6])
>3m-16y	17/63 (27.0 [16.6-39.7])	3/14 (21.4 [4/7-50.8])	0.999	20/77 (26.0 [16.6-37.2])
Over 16y	48/90 (53.3 [42.5-63.9])	7/69 (39.1 [27.6-51.6])	0.081	75/159 (47.2 [39.2-55.2])

Note:
m=months, y=years
p-value calculated using univariable logistic regression with 3GC-R alone as the predictor

The Kaplan-Meier estimation of survival function for all participants, is shown in Figure 4.3. The median survival time was 38 days for 3GC-R infection and 116 days for 3GC-S infection. There was a rapid decline in survival in the first 2 weeks following admission, which then declined to the end of the 180-day study period. Stratifying by 3GC-R status showed that this early mortality occurred at similar rates in patients with 3GC-R and 3GC-S infection. Subsequent divergence of the curves suggests a higher probability of survival in patients with 3GC-S BSI, following these early deaths, but there was no overall difference in survival on log-rank testing (p=0.34).

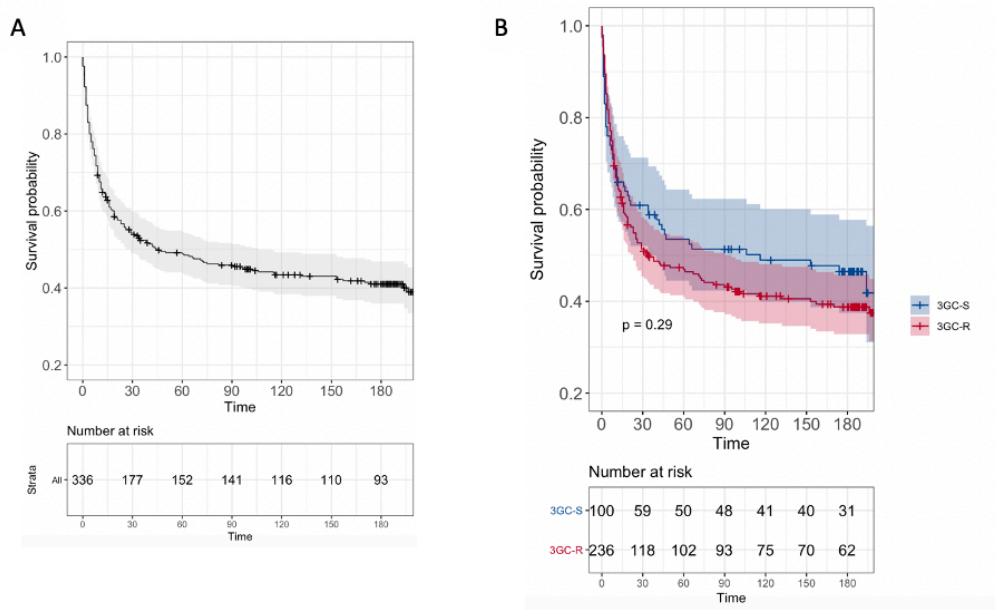


Figure 4.3: Kaplan-Meier survival curves of (A) all included participants and (B) stratified by 3GC-R status. Crosses indicate censoring. p value is derived from log-rank test comparing survival of participants with 3GC-S and 3GC-R infection.

Figure 4.4 shows the Kaplan-Meier survival curves for different age groups. Again, these stratified survival estimates are limited by smaller sample sizes, but suggest lowest survivorship in the young infants (<3 months old) with no survival difference by 3GC-R status. In older children and adults, separation of the curves after early deaths, suggests lower probability of survival in patients with 3GC-R infection.

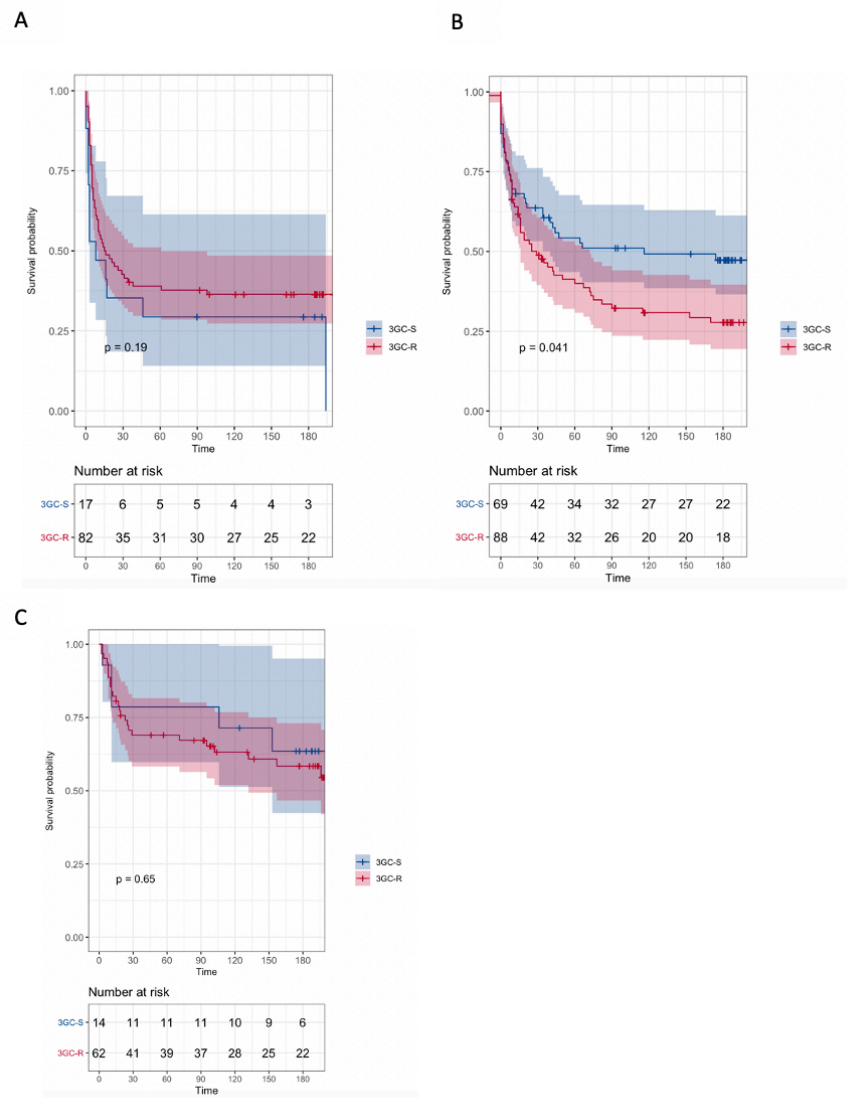


Figure 4.4: Kaplan-meier survival curves shown for (A) young infants <3months old, (B) adults > 16 years and (C) children aged over 3 months and less than 16 years. Crosses indicate censoring. p-values are derived from log-rank testing comparing survival of participants with 3GC-S and 3GC-R infection.

4.4.5 Impact of 3GC-R on in-hospital mortality and survival

The effect of 3GC-R on in-hospital mortality was estimated using a logistic regression model. Confounding variables in the models were selected a priori, as described in section 4.4.3.1. Visual inspection of the data (Table 4.2) suggested a non-linear relationship between log odds of mortality and age. This was further investigated using a cubic spline generalised additive

model (gam) to create a smoothed visualisation of this relationship (Appendix Figure 4.1).

This revealed an approximately piece-wise linear relationship, which was then replaced with a piece-wise linear term in order to avoid over-fitting.

4.4.5.1 Logistic regression model

From the baseline multivariable logistic regression model (Table 4.3, Model 1), 3GC-R was significantly associated with in-hospital mortality (total effect of 3GC-R on death) (OR 1.85, 95%CI 1.06-3.26). Older age was potentially protective against death (0.43, 95% CI 0.24-0.68). Adjustment for effective antibiotic therapy (direct effect of 3GC-R on death) (Table 4.3, Model 2) resulted in a reduction in the association of 3GC-R with death (OR 1.44, 95% CI 0.78-2.70), whilst older age remained protective.

Table 4.3: Multivariable associations with in-hospital death from logistic regression

	Model 1		Model 2	
Variable	OR (95% CI)	p-value	OR (95% CI)	p-value
Age	0.43 (0.24-0.68)	0.003	0.46 (0.26-0.72)	0.003
HIV infection	1.26 (0.61-2.60)	0.524	1.25 (0.61-2.60)	0.533
HIV exposure	0.41 (0.15-1.05)	0.070	0.43 (0.16-1.11)	0.085
3GC-R	1.85 (1.06-3.26)	0.031	1.44 (0.78-2.70)	0.247
Prior hospitalisation	1.01 (0.52-1.99)	0.962	1.03 (0.52-2.03)	0.932
Prior operation	1.00 (0.54-1.85)	0.989	1.95 (0.57-1.94)	0.879
Effective antibiotic	-	-	0.61 (0.36-1.05)	0.075

4.4.5.2 Cox proportional hazards model

Two models were again considered, a baseline model (Model 1) and a model which conditioned on use of effective antibiotics (Model 2). These models were sensitivity tested for entry point, using day of hospital admission as the model entry point instead of day of infection.
The results of this analysis were consistent with those of the primary analysis and are shown in the appendix to this chapter (Appendix Table 4.2).

3GC-R was associated with an increased probability of death in the model which did not condition on effective antibiotic use (HR 1.51, 95%CI 1.04-2.18)(Table 4.4, Model 1). After adjusting for the effective antibiotic covariate, this effect was reduced (HR 1.20 (0.80-1.80) and use of effective antibiotics was associated with increased survival probability (HR 0.62, 95%CI 0.44-0.87). Similar to the logistic regression models, increasing age was protective for mortality in both models.

Table 4.4: Multivariable associations with in-hospital death from Cox proportional hazards

	Model 1		Model 2	
Variable	HR (95% CI)	p-value	HR (95% CI)	p-value
Age	0.69 (0.51-0.95)	0.023	0.68 (0.50-0.93)	0.014
HIV	1.58 (0.97-2.57)	0.064	1.60 (0.98-2.62)	0.059
HIV exposure	0.49 (0.24-1.03)	0.059	0.50 (0.24-1.03)	0.061
3GC-R	1.51 (1.04-2.18)	0.028	1.20 (0.80-1.80)	0.400
Prior hospitalisation	0.99 (0.65-1.49)	0.944	1.19 (0.79-1.79)	0.909
Prior operation	0.89 (0.60-1.30)	0.538	0.92 (0.63-1.34)	0.660
Effective antibiotic	-	--	0.62 (0.44-0.87)	0.007

4.4.6 Length of stay

Median length of stay for participants with 3GC-S BSI was 11 days (IQR 9-15) and for participants with 3GC-R BSI was 24 days (IQR 22-31). Figure 4.5 shows the Kaplan-Meier curves of time to discharge (measured from day of admission). Curves are shown aggregated and stratified by 3GC-R status. Early separation of the curves suggests that participants with 3GC-R infection required longer hospital stays, which was significant on log-rank testing ($p<0.0001$).

.

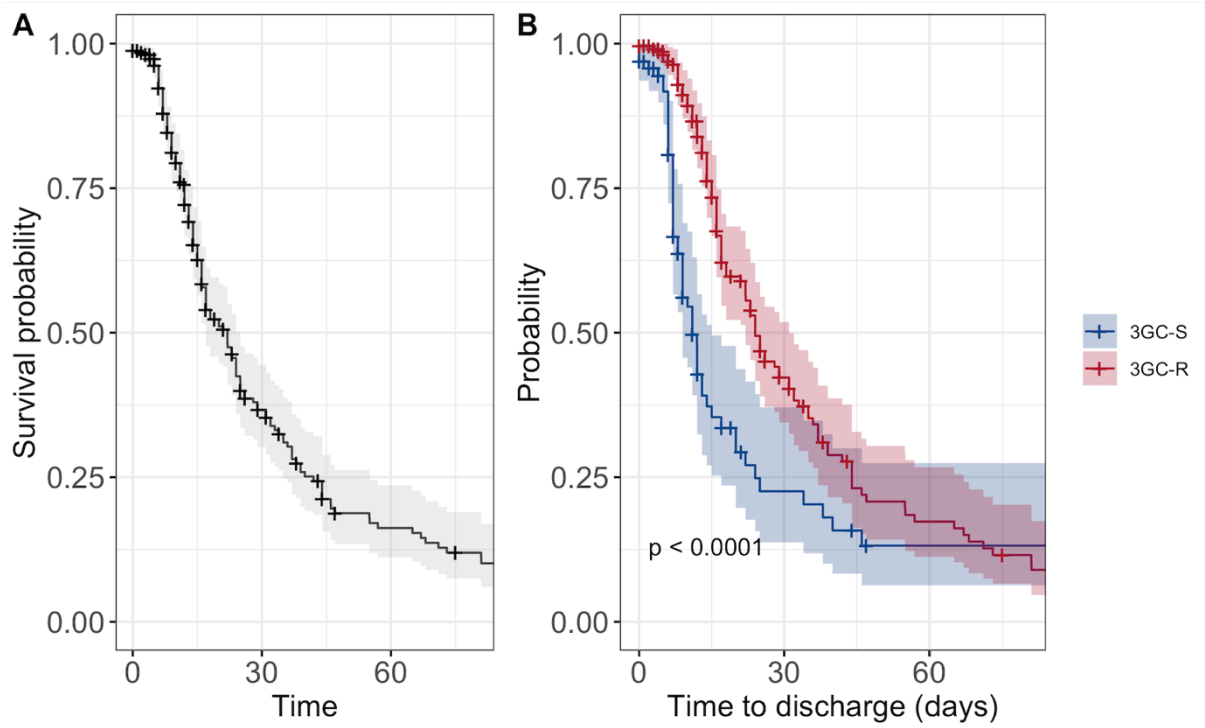


Figure 4.5: Kaplan-Meier curves showing time to discharge (A) for all included participants and (B) stratified by 3GC-R status. Crosses indicate censoring. p value is derived from log-rank test comparing length of stay for participants with 3GC-S and 3GC-R infection.

The results of cox proportional hazards models for length of stay are shown in Table 4.5. The time scale considered was time of admission to time of discharge and in the primary analysis, deaths before discharge were right-censored. Two models were again constructed, the first baseline model (Model 1) and the second (Model 2) which additionally conditioned on use of effective antibiotics. In Model 1, 3GC-R was associated with a decreased probability of hospital discharge (HR 0.70, 95%CI 0.49-0.99). This was negligible change in Model 2 (HR 0.71, 95% CI 0.50-1.01). Prior hospitalisation and prior operations were also associated with reduced probability of discharge in both models (Table 4.4). Older age was associated with increased probability of discharge, though with a small effect size (HR 1.01, 95% CI 1.00-1.02).

A sensitivity analysis was conducted in which 'death' was included as a separate covariate, to establish whether or not censoring of patients who died was introducing significant bias into the model. The output of the sensitivity analysis was similar to the primary model, suggesting that censoring deaths has minimal impact on the final results.

Table 4.5: Multivariable associations with hospital discharge from Cox proportional hazards

Variable	Model 1 HR (95% CI)	p-value	Model 2 HR (95% CI)	p-value
Age	1.01 (1.00-1.02)	0.003	1.01 (1.00-1.02)	0.002
HIV	1.30 (0.91-1.86)	0.155	1.29 (0.90-1.851)	0.172
HIV exposure	1.19 (0.63-2.25)	0.590	1.20 (0.64-2.29)	0.565
3GC-R	0.70 (0.49-0.99)	0.050	0.71 (0.50-1.01)	0.060
Prior hospitalisation	0.56 (0.37-0.84)	0.005	0.55 (0.36-0.84)	0.005
Prior operation	0.53 (0.36-0.79)	0.001	0.53 (0.36-0.79)	0.001
Effective antibiotic	-	-	1.08 (0.78-1.51)	0.643

4.5 Discussion

4.5.1 The morbidity and mortality burden from 3GC-R BSI in Blantyre, Malawi.

In this chapter I have presented the mortality and morbidity outcomes for patients who have bloodstream infection with Enterobacterales or *Acinetobacter* spp., in Blantyre. The data show that patients who have BSI caused by 3GC-R organisms, have increased risk of death and a longer hospital stay, when compared with patients who have BSI caused by 3GC-S organisms. As the first description of the clinical burden of 3GC-R BSI in Malawi, and one of few such prospective studies from sSA[16], these data make a key contribution to our understanding of AMR in sSA and provide an essential platform for future AMR research.

Mortality in these patients is extremely high, regardless of 3GC-R status, with an in-hospital case-fatality rate of 46.4% (95%CI, 41.0-51.9). Lack of similar large-scale data from sSA makes comparisons with other countries difficult, but this is substantially higher than European estimates, which are generally reported at less than 20%, even when 3GC-R is present[18]. In Malawi, where patients often have complex underlying comorbidities such as HIV and malnutrition, late presentation to secondary care is common, and delays in the acute management of sepsis are frequent, this high mortality is perhaps not unsurprising. In this study, survival was poorest in the youngest participants (<3 months old).

Although there was no difference in case-fatality rates between 3GC-R and 3GC-S BSI in these children, the study was not powered to do subgroup analysis of mortality/morbidity associations by age-group and this finding warrants potential investigation with a larger young infant or neonatal cohort.

The explanation for adverse outcomes from DRIs are likely to be multifactorial and may include both host and pathogen factors [202] [203]. However, the finding that adjustment for effective antibiotic therapy reduces the association between 3GC-R and mortality, supports the concept that poor outcomes in DRIs are in part mediated by absence of appropriate antibiotic treatments[204]. This finding has been documented elsewhere[143] and whilst perhaps unsurprising, is crucial to recognise and document in the Malawian context, where 3rd generation cephalosporins are frequently the antibiotics of first and last resort and improving access to watch and reserve antibiotics could save lives.

A strength of this study is its prospective observational design, which allowed for detailed characterisation of all BSI episodes. As a result of this, I was able to observe that, in some cases, patients whose blood culture was positive for 3GCR-E survived without effective antibiotic treatment. In this cohort, only half of participants received an antibiotic that showed in vitro activity against the bacteria isolated from their blood culture, and 70/336 (20.8%) of patients survived to 6-months having never received an effective antibiotic for their infection. Typically, a blood culture yielding a member of the family Enterobacterales would be considered to be of high clinical significance and thus trigger antimicrobial therapy to be

commenced or refined. It would be unusual to classify such an organism as a contaminant[205]. Consequently, Enterobacterales are routinely included in AMR surveillance studies without further consideration[206].

The finding that some patients survive their admission without effective treatment for apparent blood stream infection has a number of potential explanations and therefore warrants further investigation. Firstly, antimicrobial susceptibility and resistance are presented to clinicians as categories in order to rapidly distinguish which infections are likely or not to respond to antibiotics[207]. However, each isolate will have a MIC (minimum inhibitory concentration) value on a continuum, further bacterial MICs do not necessary correspond to therapeutic success or failure, particularly if they fall in an intermediate range[208] as this will depend on host factors and the nature of the infection in a given individual, including bacterial load and presence or absence of a focus of infection in addition to the bloodstream infection. Establishing ceftriaxone MICs for the cultured organisms in this cohort and corresponding these to clinical outcomes, might therefore be a useful. Secondly, Enterobacterales could be present in blood cultures as contaminants or cryptic organisms. This is commonly described in Gram-positive bacteria[209, 210], however if also true for 3GCR-E, it would have profound implications for AMR burden studies. In the final chapter of this thesis, I will describe a potential approach to classifying the role of the cultured bacteria in each patient's clinical presentation, in order to more accurately define burden.

4.5.2 Limitations

There are a number of limitations to this study which impacted the construction of the morbidity and mortality models. As discussed in section 4.4.3, the presence of co-infection with other pathogens could worsen outcomes in patients with bloodstream Enterobacterales and these data should ideally be conditioned upon in outcome models. MTB is a particularly important example, given the recent finding that one-third of septic adult inpatients at QECH had active TB, as diagnosed by urinary lipoarabinomannan (uLAM) testing (personal communication J Lewis). However, this finding was unknown at the time the current study was designed and uLAM is not yet routinely available at QECH.

Likewise, underlying host illness is a potentially important adjustment factor in models seeking to understand the direct effect of 3GC-R on outcomes. I was able to condition on some key host factors such as HIV, age and prior surgery or hospitalisation, but there are likely to be a number of unmeasured or unknown confounders in this cohort. Some studies attempt to define host health status based on scores such as the Charlson comorbidity index[131, 211], which predicts long term mortality based on multiple comorbidities such as chronic lung, liver and cardiac disease[212], but these illnesses are difficult to diagnose and define in Malawi and such scores are likely to be unreliable in this setting.

Similarly, I had insufficient data to include a sepsis severity variable as an adjustment factor in the models and this is particularly challenging in a cohort which includes both adults and children. As described above, several of the domains of sepsis scores require parameters such as blood

tests which are not available at QECH, or requirement for ventilation and inotropic support, which is not routinely assessed[213]. By recruiting participants at the point of a positive blood culture result and not at the point of presentation to hospital, I was additionally limited in my ability to collect these data as part of the study. Although future research should perhaps attempt to incorporate sepsis metrics into burden models, the validity of sepsis severity scores for predicting poor outcomes in LMICs, remains in question. In adults, studies which aimed to validate the qSOFA score in sSA, found it limited in its ability to identify those at risk of poor outcomes[214-216] and although SOFA and qSOFA criteria show some promise in identifying children at higher risk of mortality from sepsis[217, 218], there are again insufficient data from paediatric cohorts in LMICs.

Finally, this cohort incorporates a considerable amount of heterogeneity, both in terms of patients, who range from neonates to adults, and bacterial organisms which include *E. coli*, *Klebsiella* spp. and *Acinetobacter* spp. The study was not powered to enable subgroup analysis based on age or organism and whilst this does not prohibit an overall estimate of the impact of 3GC-R on outcomes, larger studies or long-term prospective surveillance might consider this approach.

4.6 Conclusions and future work

Despite these limitations, I was able to generate robust mortality and morbidity models which were consistent in the finding that third-generation cephalosporin bloodstream infections are associated with adverse clinical outcomes for patients. This study describes the first

morbidity and mortality burden estimates from 3GC-R BSI in Malawi and will therefore contribute to emerging knowledge on the burden of AMR in sub-Saharan Africa and provide impetus for future research into AMR mitigation strategies.

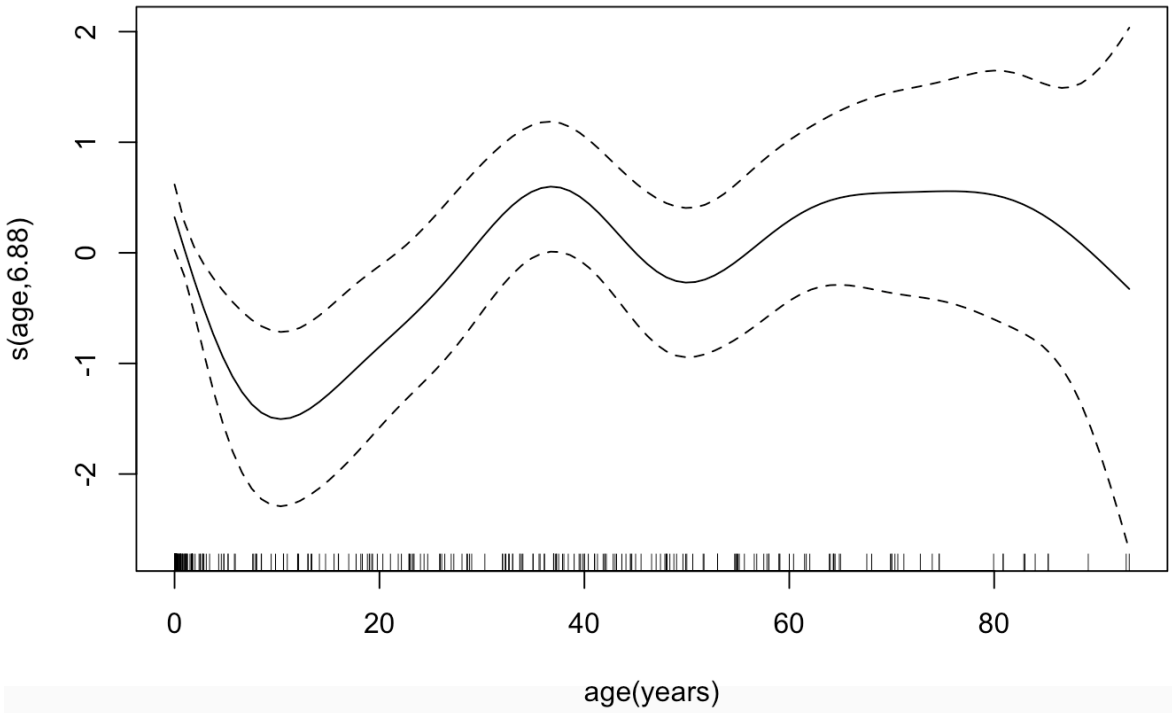
Future work should aim to elucidate the relative impact of AMR on specific cohorts such as neonates and young children and should consider the role of sepsis metrics and comprehensive microbiological diagnostics to inform clinical burden models. Incorporating health burden estimates with incidence data will additionally contribute to global burden of disease estimates.

Whilst clinical burden estimates are important, quantifying the economic outcomes of drug-resistant infections will help guide policy makers to set priorities for treatment and prevention. In the following chapter, I explore the economic burden of 3GC-R BSI in Malawi from a healthcare provider and patient perspective.

4.7 Appendix

Appendix Table 4.1: Parameters used for age-adapted SIRS scores[217].

Age group	Heat rate (min^{-1})	Respiratory rate (min^{-1})	White cell count ($x10^9$ L^{-1})	Temperature (°C)
<2 years	<90 or >180	>34	>17.5 or >5	<36.0 or >38.5
2 to 5 yrs	>140	>22	>15.5 or <6	<36.0 or >38.5
>5 to 12 yrs	>130	>18	>13.5 or <4.5	<36.0 or >38.5
>12 to <18 yrs	>110	>14	>11 or <4.5	<36.0 or >38.5



Appendix Figure 4.1: Output plot of generalised additive model (gam) showing non-linear relationship between predictor variable (age) and outcome (mortality). Death first decreases with age until around 10 years, the increases until around 38 years, then decreases to around 90 years.

Appendix Table 4.2: Multivariable associations with hospital discharge from Cox proportional hazards using day of admission as entry point

Variable	Model 1 HR (95% CI)	p-value	Model 2 HR (95% CI)	p-value
Age	1.00 (0.99-1.01)	0.780	1.00 (0.99-1.01)	0.825
HIV	1.27 (0.89-1.81)	0.180	1.31 (0.93-1.86)	0.126
HIV exposure	0.60 (0.29-1.26)	0.181	0.60 (0.29-1.25)	0.173
3GC-R	1.52 (1.06-2.19)	0.024	0.71 (0.50-1.01)	0.060
Prior hospitalisation	0.77 (0.52-1.15)	0.200	0.79 (0.53-1.17)	0.243
Prior operation	0.53 (0.36-0.79)	0.001	0.84 (0.58-1.24)	0.395
Effective antibiotic	-	-	0.61 (0.43-0.87)	0.006

Chapter 5

The economic burden of third-generation cephalosporin resistant bloodstream infection in Blantyre, Malawi

5.1 Overview

This chapter describes the health economics sub-study of the hospital cohort, which aims to quantify the health provider, individual, societal and health-related quality of life (HRQoL) costs associated with 3GC-R Enterobacterales bloodstream infection and thus estimate the economic burden of these infections in Malawi.

These data are derived from the hospital cohort, the general methods for which are described in Chapter 2. Of note, economic data were not collected from children, because the required primary costing data (e.g. cost of individual health care resources such as blood tests, see Section 5.3.1.2) are not available for paediatric admissions at QECH and HRQoL tools are not currently validated in children. The data analysis for this chapter was carried out by Dr Hendramoorthy Maheswaran (HM), a health economist with experience of work in Malawi, who was closely involved in the design of the hospital study. I carried out the data collection, data cleaning and interpretation of the analysed results.
Using previously validated data collection tools, I collected data on health provider (direct medical) costs, costs incurred by patients and their

families as a result of hospitalisation (direct non-medical and indirect costs) and health-related quality of life outcomes in patients with 3GC-R and 3GC-S bloodstream infections. I first used these data to make cost comparisons between 3GC-R and 3GC-S infections and then applied the generated costs to microbiological surveillance and projected population data for Blantyre and Malawi, to estimate the economic burden of 3GC-R BSI.

I show that 3GC-R was associated with higher health provider and patient level costs than 3GC-S infection, as well as poorer HRQoL outcomes. I demonstrate a substantial current and future economic burden to the hospital and society as a result of 3GC-R *E. coli* and *Klebsiella* spp. BSI.

5.2 Background and chapter aims

5.2.1 Summary of health economic approach

In Malawi, public health services are provided freely to residents and medical care at QECH is free of charge. Nonetheless, medical admissions incur costs to the health provider and to the patient and I estimated these costs, in relation to 3GC-R BSI, as follows. Firstly, I estimated the direct medical costs, which represent the costs incurred by QECH, in providing medical care to patients. Secondly, I estimated the costs directly incurred by patients and their guardians in relation to a hospital admission. These costs commonly include transport, food and other non-medical items and are referred to as direct non-medical costs. Thirdly, I estimated the indirect costs incurred by patients and their careers who may take time off work as a result of the hospital admission. The total of these cost categories (direct medical, direct non-medical and indirect costs) is

considered the societal cost[136]. Finally, this chapter investigates the impact of 3GC-R BSI on an individual's health-related quality of life (HRQoL).

The health economic data generated were then used in conjunction with historical microbiology surveillance data from QECH/MLW[3], to estimate the economic burden posed by 3GC-R BSI in Malawi. These burden estimates focussed on the two most commonly isolated bacteria in this cohort, (*E. coli* and *Klebsiella* spp.) and were quantified in direct medical costs, societal costs and quality-adjusted life years (QALYs) lost.

5.2.2 Chapter aims

The specific aims of this chapter are:

1. To estimate the healthcare provider costs of providing inpatient medical care to adult patients with 3GC-R BSI to QECH (direct medical costs).
2. To estimate the costs to individuals who are admitted to QECH with 3GC-R BSI (direct non-medical and indirect costs).
3. To estimate the impact of 3GC-R BSI on the health-related quality of life (HRQoL) of adult patients admitted to QECH.
4. To explore the economic burden posed by 3GC-R BSI

5.3 Methods

5.3.1 Data Collection

5.3.1.1 Overview

Health economic data were collected from adult participants of the hospital cohort study, which is described in detail in Chapter 2. The health economic methods are summarised in Figure 5.1 and the CRFs used for data collection are included as extended data in the published protocol[145]. These CRFs were developed and utilised by HM for the previous inpatient costing study at QECH[129]. All CRFs were translated into Chichewa as previously described[129].

For the sub-study, data on all health care resources used during each participant's hospital admission were extracted. Previously calculated unit costs for each health care resource use items were used, to estimate the total direct medical costs[129]. Data were also collected from participants, to capture any direct non-medical costs (e.g. food) and indirect costs (e.g. loss of income) that they or their guardian incurred as a result of the hospital admission. The approved Chichewa version of the EuroQoL EQ-5D-3L tool (hereafter EQ-5D)[151] was used to evaluate HRQoL.

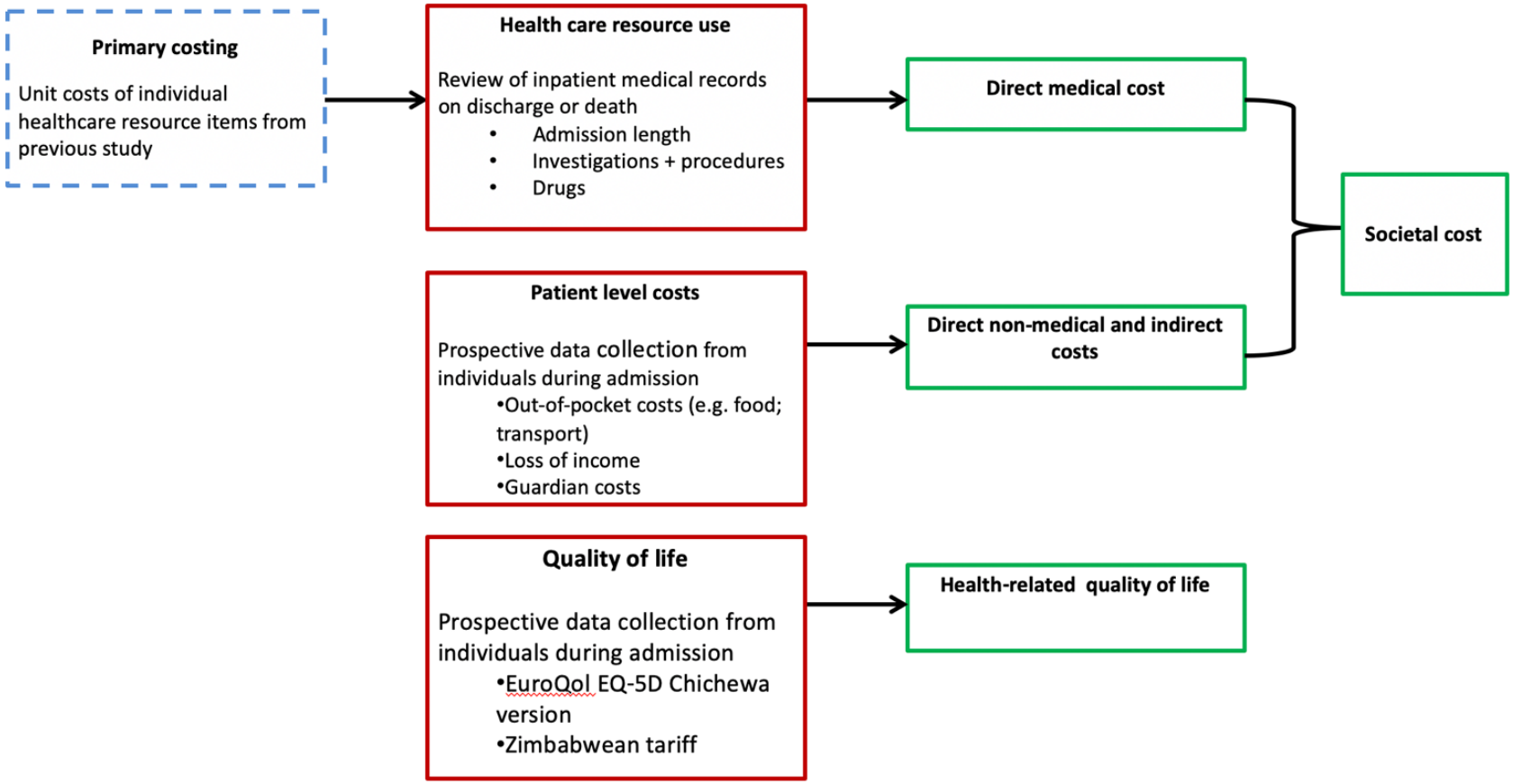


Figure 5.1: Overview of data collection and health economic analysis. Blue dashed box indicates data collected from previous study. Red boxes describe raw data collection and green boxes describe health economic outputs.

5.3.1.2 Primary costing study

A primary costing study was beyond the scope of my PhD. I utilised data from a primary hospital costing study undertaken at QECH in 2014[129]. This study generated unit costs for a range of healthcare resource-use items from the provider prospective and to my knowledge are the most recent cost estimates for inpatient healthcare resources at QECH. This previous study did not estimate unit costs for health care resources used to deliver paediatric services.

The methods for this hospital costing study are described in detail elsewhere[129]. Briefly, UNAIDS costing guidelines were used to estimate the unit costs for a wide range of healthcare resources commonly used to deliver care to medical inpatients at QECH[147]. These resources were identified via a series of interviews with health and non-health professionals working on the wards and in laboratories, radiology, pharmacy and central administration. Resources identified included staff, consumables, equipment and other overheads (e.g. utilities). These resources were then converted into unit costs using financial data from hospital administration systems, and national and international market prices. The costs of medications in my study were obtained from the International Medical Products Price Guide published by Management Sciences for Health[219].

For my study, all costs were adjusted using data reported by the World Bank and reported in 2019 US Dollars[220].

5.3.1.3 Health care resource use

To determine the resources used for each medical admission in my study, the medical notes were reviewed upon discharge or death of the participant. Information from the notes was extracted by a study clinician, to establish medications given, duration of hospital admission, types and numbers of investigations and procedures performed[145]. For medications, including intravenous fluids, the study clinician recorded the route of administration, dosage and number of doses given.

5.3.1.4 Patient costs (direct non-medical and indirect):

A previously developed questionnaire[117, 145] was used to collect data to estimate the direct non-medical and indirect costs incurred by each participant and their guardian during the hospital admission[129]. This questionnaire was administered to patients and their guardians as close to discharge as possible and ideally on the day of discharge, if this could be anticipated. Data collected included cost of transportation, food, drinks, toiletries, clothing and other items bought during the hospital admission. The total of all these costs equated to the direct non-medical cost.

For indirect costs, the time off work taken by participants or their guardians (in days), was multiplied by their self-reported daily income [221]. These direct non-medical costs and indirect costs were estimated in Malawian Kwacha (MWK); and converted to 2019 US Dollars using the prevailing exchange rate at time the analysis was undertaken (1US$=750 MWK).

5.3.1.5 Health-related quality of life (HRQoL)

The Chichewa version of the EuroQoL EQ-5D-3L was used to capture the HRQoL of participants recruited into this study[222]. EQ-5D has a descriptive component and a visual analogue scale (VAS)[145].The descriptive component assesses HRQoL across five domains: anxiety, pain, self-care, usual activities and mobility. Participants are asked to rate themselves on a 3-point ordinal scale (no problems, moderate or extreme problems). Responses to the measure would therefore generate 243 unique health states, with each one then converted into a EQ-5D utility score using a tariff set. Tariff sets have been derived in several countries through national surveys on the general population. Since no tariff set currently exists for Malawi, the Zimbabwean tariff set was used to calculate a utility score for each participant[223]. This results in EQ-5D utility scores ranging from 1.0 (representing perfect health) to -0.29. Negative EQ-5D utility scores equate to health states the general population considers worse than death. As a sensitivity analysis, the UK tariff set was also used to generate EQ-5D utility scores[224].

The visual analogue scale is similar to a thermometer, and ranges from 100 (best imaginable health state) to 0 (worst imaginable health state). Participants record how good or bad their health is on that day by drawing a line on the scale.

Of note, there are currently no HR-QoL tools which are validated for use in children and the high proportion of neonates in my main cohort, would, in any case, negate the use of these tools for the paediatric patients. The

HR-QoL data were analysed for inpatient stays only as insufficient follow-up data were collected.

5.3.2 Statistical analysis

Analysis was carried out in Stata version 13 (Stata Corporation, Texas, USA). Figures were generated using R.

The total direct medical cost per participant was estimated by summing the cost of hospital ward stay, the cost of all investigations and procedures and the cost of all medications given. The cost of hospital stay was calculated by multiplying the daily cost of admission by length of admission in days. The costs of investigations and procedures was estimated by multiplying the unit cost of each investigation/procedure by the number of times it was performed during the hospital stay. The cost of all medications given was estimated by multiplying the cost of each individual drug by the number of doses used.

The total direct non-medical and indirect costs per participant was estimated by summing the costs incurred by the participant and their guardian. The total societal cost per participant cost was estimated by adding the total direct medical cost, and the total direct non-medical and indirect cost.

EQ-5D utility and VAS scores were calculated for each participant. For the primary analysis, I present the EQ-5D utility scores generated using the Zimbabwean tariff. For the sensitivity analysis I present the EQ-5D utility scores generated using the UK tariff.

Mean total direct medical costs, direct non-medical and indirect costs and EQ-5D utility scores were calculated and stratified by the organism causing the BSI and by 3GC-R status. Means and standard errors are presented for all costs and cost differences are presented as means with 95% credible intervals (95% CR).

Multivariable analysis was then undertaken to explore the independent effects of 3GC-R on total direct medical cost, total direct non-medical and indirect cost and HRQoL outcomes. For each multivariable analysis, two alternate models were constructed. The first was adjusted for organism, age and sex. The second model was additionally adjusted for HIV status as previous work found that HIV infection was associated with higher costs of healthcare at QECH[129].

As cost and EQ-5D utility data were non-normally distributed, non-parametric bootstrapped methods were used to generate standard errors and 95% credible intervals[225].

5.3.3 Estimating the economic burden

Analysis was carried out in Microsoft Excel.

To explore the economic burden of 3GC-R in Malawi, the annual direct medical cost, annual societal cost and annual quality-adjusted life years (QALYs) lost due to *E. coli* and *Klebsiella* BSI were estimated. Estimates were made for both 3GC-R and 3GC-S BSI, from 1998 to 2030. Burden estimates were made for hospitalised adults and children.

The microbiology surveillance data for Blantyre (available for 1998-2016 for adults and children)[3] was used in conjunction with historical and projected population data for Blantyre and Malawi [226], to estimate the annual number of cases for 3GC-R and 3GC-S *E. coli* and *Klebsiella* BSI. Appendix Table 5.1 shows the total population for Blantyre and Malawi obtained from Malawi National Statistics Office[226] and Appendix Table 5.2 shows the observed and estimated annual cases of 3GC-S and 3GC-R *E. coli* and *Klebsiella* BSI in Blantyre and Malawi.

In order to make projections, a number of assumptions were made. First, it was assumed that the incidence of BSI in Blantyre reflects the incidence in Malawi as a whole. Second, to make projections beyond 2016, it was assumed that the incidence of infections remained constant at 2016 levels. Third, the direct medical, societal costs and deterioration in HRQoL associated with hospitalisation for BSI in children was assumed to be equivalent to that estimated for the adults.

To estimate the total annual direct medical and societal costs, the estimated cost per case of BSI obtained from the economic outputs as described in Section 5.3.1, was multiplied by the number of estimated annual cases.

To estimate the annual QALYs lost, the total QALYs lost amongst those who died during their hospitalisation were added to the total QALYs lost during the acute illness in survivors. To estimate the QALYs lost amongst those who died, mortality rates observed in the hospital cohort were used.

I modelled QALYs lost at the population level (including adults and children), so assumed a death from BSI resulted in a loss of 45 QALYs, that BSI cases were equally distributed amongst children and adults, and that mortality rates between children and adults were comparable. To estimate the QALYs lost amongst those who survived, the EQ-5D utility scores described in Section 5.3.1 above were used, and this detriment in HRQoL was assumed to lasts for approximately one month. The estimation of total annual QALYS lost will use the annual estimated cases for the *E. coli* and *Klebsiella* spp., as shown in Appendix Table 5.2

5.4 Results

5.4.1 Participant characteristics

Costing data were estimated for inpatient stays only. 154 out of 159 adult participants were included in the health economic sub-study with the remaining five adults not included due to missing data. Direct medical cost data were available for 127/154 participants with the remainder not included due to missing patient files at the time of discharge or death. Direct non-medical and indirect cost data were available for 74/154 participants with the remainder missing due to patients dying before questionnaires could be administered. EQ-5D utility scores were available for 106/154 participants and EQ-5D VAS scores were available for 100/154 participants. Again, incomplete HRQoL data sets are due to patients dying before data could be collected.

Table 5.1 shows the characteristics of included participants. There was no significant difference in key characteristics between adults in the full cohort

and those included in the economic sub-study. Median age of participants was 45.2 years (48.2 for adults in the full cohort). 55.8% of participants were HIV infected (58.7% in the full cohort) and 84.9% of HIV infected adults were on ART (81.3% in full cohort).

The median duration of hospital stay was 9.0 days (IQR 4.8-16.0) and the in-hospital mortality was 72/154 46.8%.

E. coli and *Klebsiella* spp. were the most frequently isolated organisms in the sub-study, as in the full cohort, causing 69.5% and 15.6% of BSI respectively. Just over half (55.2%) of organisms were 3GC-R.

Table 5.1: Participant characteristics

	Characteristic	n=154
Age	Median (IQR)	45.4 (42.5, 48.4)
Sex	Female	78 (50.6%)
	Male	76 (49.4%)
Education	No Formal Schooling	3 (2.0%)
	Any Primary	56 (36.4%)
	Any Secondary	34 (22.1%)
	College or higher	21 (13.6%)
	Not Known	40 (26.0%)
Employment	Paid employee	18 (12.3%)
	Paid domestic worker	1 (0.7%)
	Self-employed	29 (19.9%)
	Unemployed	52 (35.6%)
	Student	6 (4.1%)
	Other	40 (27.4%)
HIV status	Negative	64 (41.6%)
	Positive	86 (55.8%)
	Unknown	4 (2.6%)
ART status	Not on ART	13 (15.1%)
	On ART	73 (84.9%)
Organism	*E. coli*	107 (69.5%)
	Klebsiella spp.	24 (15.6%)
	Other*	23 (14.9%)
3GC status	3GC-S	69 (44.8%)
	3GC-R	85 (55.2%)

Note:

* Other organisms: *Proteus mirabilis* and *Enterobacter* spp.

5.4.2 Direct medical costs

Table 5.2 shows the mean total direct medical costs for participants in the economic sub-study, stratified by organism. Table 5.3 shows these costs stratified by 3GC susceptibility status, as well as the mean difference in costs between those who had 3GC-R and 3GC-S infections.

The mean total direct medical cost across all 127 participants was US$289.67 (SE: 22.8). The majority of these costs were accounted for by ward stay (US$155.25, SE:17.5) and investigations (US$98.84, SE: 5.7) (Table 5.1). The mean total direct medical cost of an admission with *Klebsiella* BSI was higher than for *E. coli* at US$359.29 (SE: 68.7) vs US$260.80 (SE: 26.0) (Table 5.2).

The mean total direct medical cost of an admission with any 3GC-R organism was US$334.92 (SE: 32.5) and with any 3GC-S organism was US$228.50 (SE: 25.0). The mean total direct medical cost was US$106.42 (95%CR; 23.82-189.01) higher amongst those who had a resistant BSI than those admitted with a sensitive BSI. For each resource-use category (ward stay; medications; investigations; procedures), the direct medical costs were higher for those who had 3GC-R organism than those who had 3GC-S organism (Table 5.2).

Table 5.2: Direct medical costs by causative organism and resource-use category.

	N	2019 US Dollars Mean/SE	% of Total cost Mean*
E. coli			
Ward stay		132.01 (18.0)	43.1
Medications		27.49 (7.6)	7.7
Investigations	85	94.96 (6.9)	46.4
Procedures		6.34 (1.1)	2.9
Total		**260.80 (26.0)**	-
Klebsiella spp.			
Ward stay		217.68 (60.9)	54.3
Medications		21.00 (7.9)	6.7
Investigations	23	88.73 (8.7)	32.8
Procedures		21.88 (22.6)	6.3
Total		**359.29 (68.7)**	-
Other**			
Ward stay		183.64 (40.9)	47.0
Medications		14.85 (2.3)	5.6
Investigations	19	128.46 (19.5)	45.0
Procedures		7.63 (2.7)	2.4
Total		**334.57 (49.9)**	-
All Organisms			
Ward stay		155.25 (17.5)	45.7
Medications		24.42 (5.3)	7.2
Investigations	127	98.84 (5.7)	43.7
Procedures		11.16 (4.2)	3.4
Total		**289.67 (22.8)**	-

Note:

*Mean of percentage at the participant level

** Other organisms: *Proteus mirabilis* and *Enterobacter* spp.

Table 5.3**:** Direct medical costs by 3GC susceptibility status and resource-use category.

	n	2019 US Dollars Mean/SE	% of Total cost Mean*	Mean Differences: 3GC-R v 3GC-S (95% CR)**
3GC-S				
Ward stay		119.17 (22.1)	44.5	
Medications		11.53 (1.2)	6.0	
Investigations	54	94.05 (9.4)	47.7	-
Procedures		3.75 (0.8)	1.8	
Total		228.50 (29.0)	-	
3GC-R				
Ward stay		181.93 (25.4)	46.6	62.75 (-0.97, 126.48)
Medications		33.96 (9.1)	8.0	23.48 (5.99, 40.97)
Investigations	73	102.39 (7.3)	40.7	8.34 (-13.59, 30.26)
Procedures		16.64 (7.2)	4.6	13.06 (-1.07, 27.20)
Total		334.92 (32.5)	-	107.64 (25.74, 189.53)

Note:

*Mean of percentage at the participant level

**Bootstrapped estimates of mean differences and 95% credible interval (95% CR)

5.4.3 Direct non-medical and indirect costs

Tables 5.4 shows the mean total direct non-medical and indirect costs, stratified by organism. The costs are shown broken down by those incurred by the patient and those incurred by the patient's guardian. Table 5.5 shows these costs stratified by 3GC susceptibility status and the mean difference in costs between those who had 3GC-R and 3GC-S infections.

The mean total direct non-medical cost associated with an admission was US$121.34 (SE: 15.4) and the mean total indirect cost was US$157.39 (SE: 55.3). The mean total direct non-medical cost of an admission with *E. coli* was US$121.27 (18.3) and for *Klebsiella* was US$71.28 (20.4) The mean

total indirect cost of an admission with *E. coli* was US$117.16 (61.7) and for *Klebsiella* was US$85.67 (43.9) (Table 5.4).

The mean total direct non-medical cost of an admission with any 3GC-R organism wasUS$134.97 (SE: 23.5) and with any 3GC-S organism was US$104.82 (SE: 18.5). In comparison to those admitted with a sensitive BSI, those who had a resistant BSI incurred a mean additional direct non-medical cost of US$30.16 (95%CR; -27.24, 87.55).

The mean total indirect cost of an admission with any 3GC-R organism was US$232.03 (SE: 97.3) and with any 3GC-S organism was US$64.67 (SE: 21.2). In comparison to those admitted with a sensitive BSI, those who had a resistant BSI incurred a mean additional indirect cost of US$167.36 (95%CR; -31.88, 366.60) (Table 5.5).

Table 5.4: Direct non-medical and indirect costs incurred by patient and family/carer by causative organism.

	N	2019 US Dollars Mean/SE
E. coli		
Patient direct non-medical		45.69 (4.8)
Patient indirect		37.81 (13.8)
Family/carer direct non-medical	51	75.58 (15.4)
Family/carer indirect		79.35 (60.6)
Total direct non-medical		**121.27 (18.3)**
Total indirect		**117.16 (61.7)**
Klebsiella		
Patient direct non-medical		31.32 (9.0)
Patient indirect		67.89 (43.8)
Family/carer direct non-medical	12	39.96 (13.2)
Family/carer indirect		17.78 (11.1)
Total direct non-medical		**71.28 (20.4)**
Total indirect		**85.67 (43.9)**
Other*		
Patient direct non-medical		63.20 (20.1)
Patient indirect		181.22 (117.4)
Family/carer direct non-medical	11	114.35 (44.2)
Family/carer indirect		240.97 (170.1)
Total direct non-medical		177.55 (52.9)
Total indirect		422.19 (226.4)
All Organisms		
Patient direct non-medical		45.96 (4.7)
Patient indirect		64.01 (21.3)
Family/carer direct non-medical	74	75.57 (12.7)
Family/carer indirect		93.39 (48.8)
Total direct non-medical		**121.53 (15.4)**
Total indirect		**157.39 (55.3)**

Note:

*Other organisms: *Proteus mirabilis* and *Enterobacter* spp

Table 5.5: Direct non-medical and indirect costs incurred by patient and family/carer by 3GC susceptibility status.

	N	2019 US Dollars Mean/SE	Mean Differences: 3GC-R v 3GC-S (95% CR)*
3GC-S			
Patient direct non-medical		40.79 (6.3)	
Patient indirect		49.42 (20.2)	
Family/carer direct non-medical	33	64.03 (14.7)	-
Family/carer indirect		15.26 (5.7)	
Total direct non-medical		**104.82 (18.5)**	
Total indirect		**64.67 (21.2)**	
3GC-R			
Patient direct non-medical		50.12 (6.8)	9.33 (-8.85, 27.52)
Patient indirect		75.75 (34.9)	26.33 (-52.88, 105.54)
Family/carer direct non-medical	41	84.85 (19.7)	20.82 (-26.66, 68.30)
Family/carer indirect		156.28 (87.2)	141.10 (-43.61, 316.66)
Total direct non-medical		**134.97 (23.5)**	**30.16 (-27.24, 87.55)**
Total indirect		**232.03 (97.3)**	**167.36 (-31.88, 366.60)**

Note:
*Bootstrapped estimates of Mean differences and 95% credible interval (95% CR)

5.4.4 Total societal cost of admission

Table 5.6 shows the total societal cost of a hospital admission with Enterobacterales BSI, stratified by organism and by 3GC susceptibility status. The mean total societal cost of a hospital admission across all participants was US$625.58 (SE 77.6). The cost of an admission with *E. coli* BSI was $US551.84 (SE,86.3) and the cost of admission with *Klebsiella* was US$585.56 (SE:164.2).

The mean total societal cost of an admission with any 3GC-R organism was US$778.72 (SE: 125.3) and with any 3GC-S organism was US$435.31 (SE: 66.2). The mean total societal cost for those admitted with a resistant BSI was US$343.41 (95%CR; -76.47, 610.34) higher than those admitted with a sensitive BSI.

Table 5.6: Total societal costs, by organism and 3GC susceptibility status.

	N	2019 US Dollars Mean/SE	Mean Differences: 3GC-R v 3GC-S (95% CR)*
All	74	625.58 (77.6)	-
Organism			
E. coli	51	551.85 (86.3)	
Klebsiella	12	585.56 (164.2)	-
Other	11	1011.13 (270.1)	
3GC-S status			
Negative	33	435.31 (66.2)	-
Positive	41	778.72 (125.3)	343.41 (76.47, 610.34)

Note:

*Bootstrapped estimates of Mean differences and 95% credible interval (95% CR)

Table 5.7: Multivariable regression analysis of costs.

Organanism/ 3GC status	**Total direct medical cost (2019 US Dollars)**		**Total societal costs (2019 INT Dollars)**	
	Model 1 (n=127) Coef (95% CR)	Model 2 (n=127) Coef (95% CR)	Model 1 (n=74) Coef (95% CR)	Model 2 (n=74) Coef (95% CR)
E. Coli	Ref	Ref	Ref	Ref
Klebsiella	60.62 (-78.17, 199.41)	55.04 (-88.64, 198.72)	-87.51 (-455.49, 280.46)	-106.66 (-540.39, 327.07)
Other	33.48 (-79.42, 146.38)	-44.03 (-67.64, 155.69)	310.32 (-261.85, 882.50)	286.11 (-300.87, 873.09)
3GC-S	**Ref**	**Ref**	**Ref**	**Ref**
3GC-R	**81.19 (3.61, 164.76)**	**84.88 (-3.11, 172.88)**	**272.85 (1.59, 544.11)**	**251.04 (-12.01, 514.10)**

Note:
Model 1: additionally adjusted for age and sex
Model 2: additionally adjusted for age, sex and HIV status
CR = Credible interval, INT = International

5.4.6 Health-related quality of life outcomes

Table 5.8 shows the EQ-5D utility and VAS scores for all participants, stratified by organism. Table 5.9 shows utility and VAS scores stratified by 3GC susceptibility status.

EQ-5D utility scores generated using the Zimbabwean tariff for participants admitted with *E. coli* BSI were 0.418 (SE: 0.04) and with *Klebsiella* BSI were 0.344 (SE,0.08). Participants with 3GC-S infection had EQ-5D utility scores of 0.481 (SE: 0.05) and those with resistant infection had EQ-5D utility scores of 0.281 (SE: 0.05). Participants with resistant infections had a EQ-5D utility score that was therefore 0.200 (95% CR:

0.061, 0.339) lower than those with sensitive infections. EQ-5D utility scores generated using the UK tariff were lower, and the difference in EQ-5D utility scores between those with sensitive and resistant infections more pronounced.

VAS scores for participants admitted with *E. coli* BSI were 58.0 (SE: 2.5) and with *Klebsiella* spp were 61.6 (SE: 5.1). VAS scores were comparable in those with resistant and sensitive infection at 59.5 and 58.2 respectively, mean difference -1.3 (95% CR: -9.0, 6.4).

Table 5.8: Health-related quality of life outcomes by causative organism.

	N	Mean/SE
E. coli		
VAS score	72	58.0 (2.5)
EQ-5D utility score (Zim Tariff)	77	0.418 (0.04)
EQ-5D utility score (UK Tariff)	77	0.184 (0.06)
Klebsiella		
VAS score	14	61.6 (5.1)
EQ-5D utility score (Zim Tariff)	14	0.344 (0.08)
EQ-5D utility score (UK Tariff	14	0.050 (0.12)
Other		
VAS score	14	59.6 (4.8)
EQ-5D utility score (Zim Tariff)	15	0.215 (0.09)
EQ-5D utility score (UK Tariff	15	-0.104 (0.13)
All Organisms		
VAS score	100	58.8
EQ-5D utility score (Zim Tariff)	106	0.379 (0.04)
EQ-5D utility score (UK Tariff	106	-0.125 (0.05)

Note:
VAS = Visual analogue scale, SE = Standard error

Table 5.9: Health-related quality of life outcomes by 3GC susceptibility status.

	N	Mean/SE	Mean Differences: 3GC-R v 3GC-S (95% CR) *
3GC-S			
VAS score	47	59.5 (3.3)	-
EQ-5D utility score (Zim Tariff)	52	0.481 (0.05)	
EQ-5D utility score (UK Tariff)	52	0.271 (0.07)	
3GC-R			
VAS score	53	58.2 (2.6)	-1.3 (-9.0, 6.4)
EQ-5D utility score (Zim Tariff)	54	0.281 (0.05)	0.200 (-0.339, -0.061)
EQ-5D utility score (UK Tariff	54	-0.015 (0.17)	-0.285 (-0.489, -0.082)

Note:

*Bootstrapped estimates of Mean differences and 95%CI

CR = Credible interval, SE = Standard error, VAS = visual analogue scale

Table 5.10 shows the findings of the multivariable analysis to investigate the independent effect of 3GC-R on EQ-5D utility scores. In model 1, after adjusting for age, sex and organism, the EQ-5D utility score (generated using the Zimbabwean tariff) in those with a 3GC-R BSI was 0.188 lower than those with a 3GC-S BSI. In model 2, additionally adjusted for HIV status, the EQ-5D utility score in those with a 3GC-R BSI was 0.157 lower than those with a 3GC-S BSI. Again, the differences were more pronounced when the UK tariff was used to generate EQ-5D utility scores.

Table 5.10: Multivariable regression analysis EQ-5D utility scores.

Organism and 3GC susceptibility status	**EQ-5D utility score – Zim Tariff**		**EQ-5D utility score – UK Tariff**	
	Model 1 (n=106) Coef (95% CR)	Model 2 (n=106) Coef (95% CR)	Model 1 (n=106) Coef (95% CR)	Model 2 (n=106) Coef (95% CR)
E. coli	Ref	Ref	Ref	Ref
Klebsiella	-0.001 (-0.209, 0.207)	0.030 (-0.176, 0.237)	-0.030 (-0.332, 0.271)	0.013 (-0.288, 0.314)
Other	-0.157 (-0.376, 0.062)	-0.114 (-0.338, 0.109)	-0.223 (-0.532, 0.086)	-0.165 (-0.483, 0.152)
3GC-S	Ref	Ref	Ref	Ref
3GC-R	-0.188 (-0.330, -0.046)	-0.157 (-0.305, -0.009)	-0.263 (-0.469, -0.056)	-0.220 (-0.436, -0.003)

Note:
Model 1: adjusted for age and sex
Model 2: adjusted for age, sex and HIV status
Zim = Zimbabwe, CR = Credible interval

5.4.7 Economic burden of 3GC-R BSI

The annual direct medical and societal cost estimates for Malawi, associated with 3GC-R and 3GC-S *E. coli* BSI admissions are shown in Table 5.11. and in Figure 5.2. Data collection for the cohort took place from 2018-2020 (see Chapter 3). In 2019, it is estimated that the annual direct medical cost to provide hospital care for those admitted with *E. coli* BSI in Malawi was US$818,469 and of this, US$605,764 was from 3GC-R. The annual societal cost for hospitalisation with *E. coli* is estimated to be US$1,808,781 in 2019, of which US$1,400,764 was accounted for by 3GC-R.

The annual direct medical and societal cost estimates for Malawi, associated with 3GC-R and 3GC-S *Klebsiella* BSI admissions are shown in Table 5.12. and in Figure 5.3. For *Klebsiella spp.* it is estimated that in 2019, the annual direct medical cost was US$703,159 and of this, US$670,670 was due to 3GC-R. The annual societal cost associated with hospitalisation with *Klebsiella spp.* BSI is estimated to be US$1,177,868, in 2019, of which US$1,117,449 is accounted for by 3GC-R organisms.

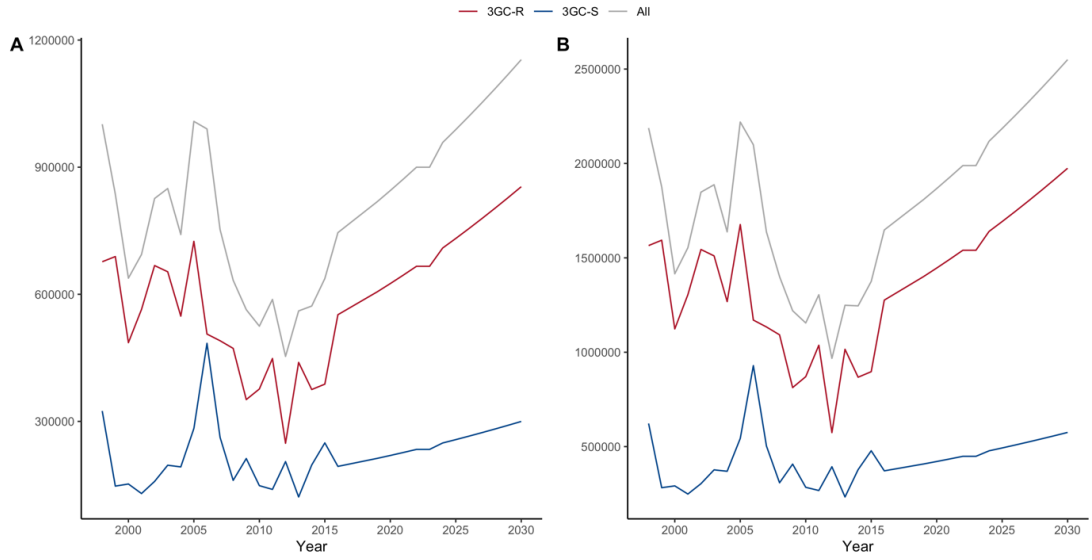


Figure 5.2: (A) Annual direct medical cost (2019 US$) and (B) Annual societal cost estimates for *E. coli* BSI between 1998 and 2030.

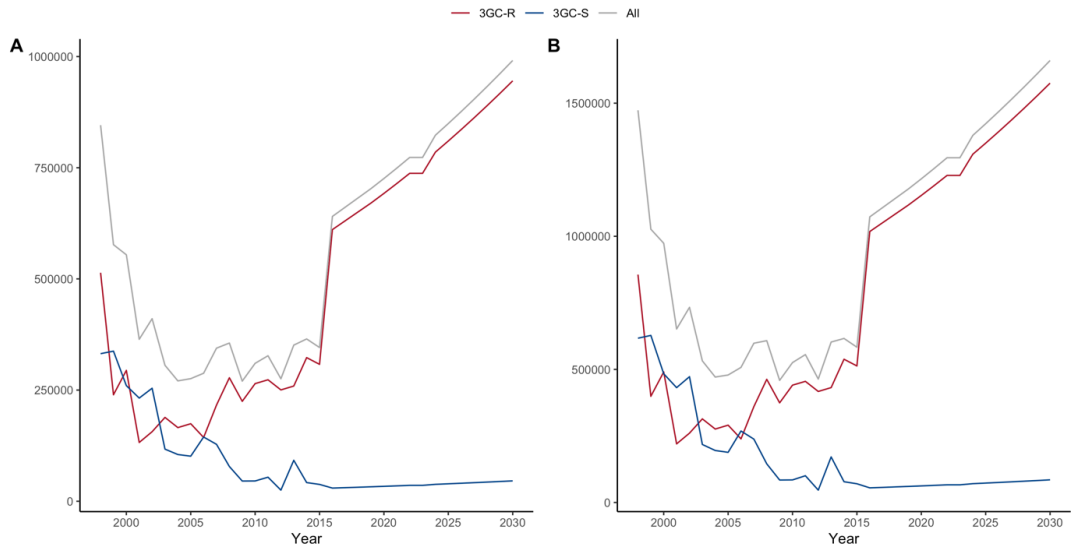


Figure 5.3: (A) Annual direct medical cost (2019 US$) and (B) Annual societal cost estimates for *Klebsiella spp.* BSI between 1998 and 2030.

The estimated total QALYs lost from *E. coli* and *Klebsiella* BSI are shown in Table 5.13 and in Figure 5.4, stratified by 3GC-R status. In 2019, it is estimated that *E. coli* infections will account for 65,548 QALYs lost, of which 50,583 are from 3GC-R. In the same time period, it is estimated that *Klebsiella* BSI will account for 41,157 QALYs lost, of which 37,207will be from 3GC-R.

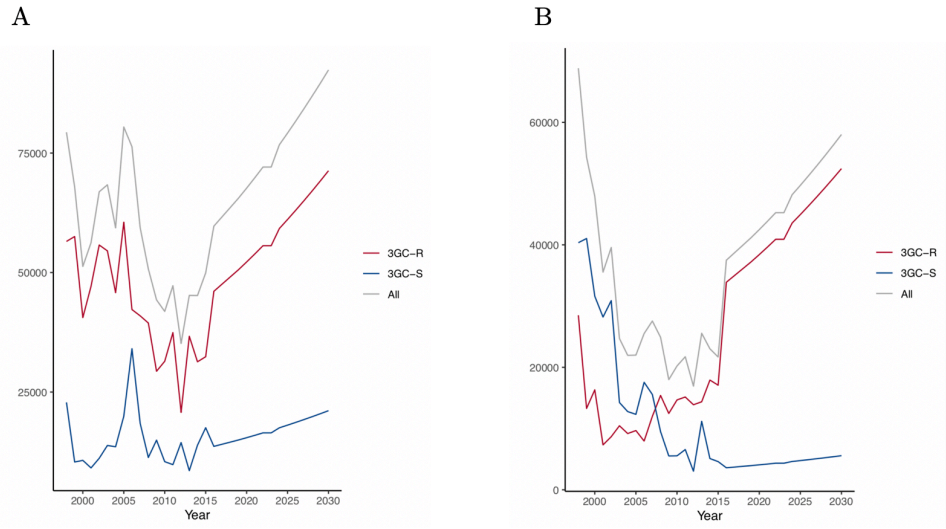


Figure 5.4: Annual QALYs lost for (A)*E. coli* and (B) *Klebsiella* BSI

Table 5.11: Annual costs (2019 US Dollars) for *E. coli*.

Year	Direct medical cost (Health Provider)			Societal cost		
	E. coli 3GC-R	*E. coli* 3GC-S	All *E. coli*	*E. coli* 3GC-R	*E. coli* 3GC-S	All *E. coli*
1998	676,798	324,590	1,001,388	1,565,020	622,642	2,187,662
1999	689,042	147,228	836,270	1,593,334	282,418	1,875,753
2000	485,881	152,210	638,091	1,123,547	291,975	1,415,522
2001	564,279	129,675	693,954	1,304,833	248,748	1,553,581
2002	667,796	158,315	826,110	1,544,204	303,685	1,847,889
2003	653,140	196,490	849,630	1,510,315	376,914	1,887,229
2004	548,452	192,580	741,032	1,268,234	369,414	1,637,648
2005	724,882	283,211	1,008,093	1,676,211	543,266	2,219,477
2006	506,089	484,199	990,288	1,170,275	928,810	2,099,084
2007	490,353	262,339	752,693	1,133,888	503,229	1,637,117
2008	472,311	160,801	633,112	1,092,167	308,454	1,400,621
2009	351,432	212,288	563,720	812,649	407,218	1,219,867
2010	376,447	148,303	524,751	870,493	284,481	1,154,974
2011	448,372	139,451	587,823	1,036,811	267,501	1,304,312
2012	248,407	205,042	453,449	574,413	393,320	967,733
2013	439,157	121,573	560,731	1,015,503	233,207	1,248,709
2014	375,185	197,075	572,260	867,575	378,037	1,245,612
2015	387,892	249,201	637,093	896,957	478,027	1,374,984
2016	551,807	193,758	745,565	1,275,992	371,674	1,647,666
2017	569,792	200,073	769,866	1,317,583	383,788	1,701,371
2018	587,778	206,389	794,167	1,359,173	395,903	1,755,076
2019	605,764	212,704	818,469	1,400,764	408,018	1,808,781
2020	625,260	219,550	844,810	1,445,844	421,149	1,866,993
2021	645,360	226,608	871,967	1,492,324	434,687	1,927,011
2022	666,053	233,874	899,926	1,540,174	448,625	1,988,799
2023	666,053	233,874	899,926	1,540,174	448,625	1,988,799
2024	709,202	249,025	958,226	1,639,951	477,689	2,117,640
2025	731,756	256,945	988,701	1,692,107	492,881	2,184,987
2026	754,955	265,090	1,020,046	1,745,752	508,506	2,254,258
2027	778,774	273,454	1,052,229	1,800,830	524,550	2,325,380
2028	803,192	282,028	1,085,220	1,857,293	540,997	2,398,290
2029	828,232	290,820	1,119,052	1,915,196	557,862	2,473,058
2030	853,926	299,842	1,153,768	1,974,610	575,169	2,549,779

Note:

Grey shading indicates data are based on projected BSI incidence (post 2016).

Table 5.12: Annual costs (2019 US Dollars) for *Klebsiella* spp..

	Direct medical cost (Health Provider)			Societal cost		
Year	*Klebsiella* 3GC-R	*Klebsiella* 3GC-S	All *Klebsiella*	*Klebsiella* 3GC-R	*Klebsiella* 3GC-S	All *Klebsiella*
1998	513,815	331,876	845,692	856,103	617,175	1,473,278
1999	239,457	337,523	576,980	398,976	627,675	1,026,651
2000	294,265	259,967	554,231	490,294	483,448	973,743
2001	132,182	232,026	364,208	220,238	431,487	651,726
2002	156,350	254,109	410,459	260,505	472,555	733,060
2003	188,507	117,192	305,699	314,084	217,936	532,020
2004	165,604	105,055	270,659	275,925	195,365	471,290
2005	174,344	101,349	275,693	290,487	188,474	478,960
2006	143,457	144,395	287,852	239,024	268,524	507,548
2007	216,217	128,018	344,235	360,254	238,069	598,323
2008	277,682	78,095	355,777	462,665	145,229	607,894
2009	224,694	45,416	270,110	374,377	84,458	458,835
2010	264,549	45,608	310,157	440,783	84,815	525,598
2011	273,146	54,062	327,208	455,108	100,537	555,645
2012	250,361	25,014	275,375	417,143	46,518	463,662
2013	259,062	92,031	351,094	431,642	171,147	602,788
2014	322,992	41,979	364,971	538,159	78,066	616,225
2015	307,868	37,859	345,726	512,959	70,404	583,363
2016	610,931	29,595	640,526	1,017,914	55,037	1,072,951
2017	630,844	30,560	661,404	1,051,092	56,831	1,107,923
2018	650,757	31,525	682,281	1,084,271	58,625	1,142,896
2019	670,670	32,489	703,159	1,117,449	60,419	1,177,868
2020	692,254	33,535	725,789	1,153,412	62,363	1,215,775
2021	714,508	34,613	749,121	1,190,491	64,368	1,254,859
2022	737,418	35,723	773,141	1,228,663	66,432	1,295,095
2023	737,418	35,723	773,141	1,228,663	66,432	1,295,095
2024	785,190	38,037	823,227	1,308,259	70,735	1,378,995
2025	810,162	39,247	849,408	1,349,866	72,985	1,422,851
2026	835,846	40,491	876,337	1,392,661	75,299	1,467,960
2027	862,217	41,768	903,986	1,436,600	77,675	1,514,274
2028	889,251	43,078	932,329	1,481,643	80,110	1,561,753
2029	916,974	44,421	961,395	1,527,834	82,608	1,610,441
2030	945,421	45,799	991,220	1,575,231	85,170	1,660,401

Note:

Grey shading indicates data are based on projected BSI incidence (post 2016).

Table 5.13: Annual QALYs lost due to *E. coli* and *Klebsiella*

Year	Total QALYs Lost			Total QALYs Lost		
	E. coli Resistant	*E. coli* Sensitive	All *E. coli*	*Klebsiella* Resistant	*Klebsiella* Sensitive	All *Klebsiella*
1998	56,514	22,838	79,352	28,505	40,351	68,856
1999	57,537	10,359	67,896	13,284	41,037	54,322
2000	40,572	10,709	51,282	16,325	31,608	47,933
2001	47,119	9,124	56,243	7,333	28,210	35,544
2002	55,763	11,139	66,901	8,674	30,895	39,569
2003	54,539	13,825	68,364	10,458	14,249	24,706
2004	45,797	13,550	59,347	9,187	12,773	21,960
2005	60,530	19,926	80,456	9,672	12,322	21,994
2006	42,260	34,068	76,327	7,959	17,556	25,515
2007	40,946	18,458	59,404	11,995	15,565	27,560
2008	39,439	11,314	50,753	15,405	9,495	24,900
2009	29,345	14,936	44,282	12,465	5,522	17,987
2010	31,434	10,434	41,869	14,676	5,545	20,222
2011	37,440	9,812	47,252	15,153	6,573	21,726
2012	20,743	14,427	35,169	13,889	3,041	16,931
2013	36,671	8,554	45,224	14,372	11,190	25,562
2014	31,329	13,866	45,195	17,919	5,104	23,023
2015	32,390	17,533	49,923	17,080	4,603	21,683
2016	46,077	13,633	59,710	33,893	3,598	37,491
2017	47,579	14,077	61,656	34,997	3,716	38,713
2018	49,081	14,521	63,602	36,102	3,833	39,935
2019	50,583	14,966	65,548	37,207	3,950	41,157
2020	52,211	15,447	67,658	38,404	4,077	42,482
2021	53,889	15,944	69,833	39,639	4,208	43,847
2022	55,617	16,455	72,072	40,910	4,343	45,253
2023	55,617	16,455	72,072	40,910	4,343	45,253
2024	59,220	17,521	76,741	43,560	4,625	48,185
2025	61,104	18,078	79,182	44,946	4,772	49,717
2026	63,041	18,651	81,692	46,370	4,923	51,293
2027	65,030	19,240	84,270	47,833	5,078	52,912
2028	67,069	19,843	86,912	49,333	5,238	54,571
2029	69,159	20,462	89,621	50,871	5,401	56,272
2030	71,305	21,097	92,401	52,449	5,568	58,018

5.5 Discussion

In this chapter, I have estimated the costs incurred by the healthcare provider, patients and their guardians, as well as the impact on the individuals' health-related quality of life, as a consequence of hospitalisation with Enterobacterales bloodstream infections in Malawi. Additionally, I have explored whether these costs and HRQoL outcomes are different amongst those admitted with 3GC-S and 3GC-R infections.

I found hospital admission with Enterobacterales BSI placed a substantial financial burden on the hospital, as well as on the patient and their families and that this burden was substantially higher amongst those who had infections that were resistant to third-generation cephalosporins. Additionally, those admitted to hospital with resistant infections had poorer HRQoL. I used the costs generated in this study, to estimate the economic burden of *E. coli* and *Klebsiella* BSI in Malawi, and found that 3GC-R accounts for more than 80% of the economic and health burden posed from these infections.

The average healthcare provider cost of managing patients with BSI in my cohort was US$289.67, comparable to estimates from the previous study of adult medical inpatients at QECH, in which the average cost of admission was US$313.65 [129]. In my study, patients admitted with resistant infections were associated with an additional US$106.42 cost to the health provider than 3GC-S infections. To put this into context, the annual cost of providing anti-retroviral treatment in Malawi is approximately US$170 [117]. AMR health provider cost data from sSA are extremely limited, with

only one other study, from Senegal, reporting costs associated with 3GC-R infections. This study found the additional cost associated with 3GC-R was 100EUR (US$120.96) but did not report mean overall costs [103].

Ward stay and investigations accounted for the majority of the total direct medical costs of 3GC-R and 3GC-R infections. Only 7.2 % of total costs were accounted for by spending on drugs, which is likely to reflect the WHO prequalification of medicines programme, which ensures availability of high quality medicines in Africa at reasonable prices. The costs of medications used in my study reflect the median costs paid by ministries of health in resource-constrained predominantly African countries [219]. In the Senegalese study and in high income settings, the majority of the additional costs of 3GC-R infections come from the use of more expensive second and third-line antibiotics, such as carbapenems and aminoglycosides [103]. In my study, there was no difference in the proportion spent on medicines between resistant and sensitive infections, likely because these more expensive antibiotics are not yet routinely available at QECH. However, as access to these drugs improves in Malawi, the cost of managing 3GC-R will rise even further.

Medical care at QECH is free of charge, but patients and their guardians inevitably incur some out-of-pocket costs as a result of a hospital admission, including for food and transportation. For patients and guardians together, the average spending on these non-medical items was US$121.53 and there was no substantial difference between 3GC-R and 3GC-S infection (mean difference US$30.16). However, this out-of-pocket cost is substantial considering that the majority of Malawians live on less

than US$2 per day [227] and hospitalisation clearly poses a significant burden on household finances, potentially pushing those affected further into poverty. This is further compounded if we additionally consider the impact of these infection on household incomes. The average indirect costs for families, incurred from loss of income was US$157.39 and this was significantly higher in 3GC-R than 3GC-S infection, with a mean difference of US$167.36. The impact of these admissions on household finances is potentially devastating and savings for households from reducing resistant infections alone would be considerable.

The health-related quality of life amongst patients in the study was poor, with an average overall EQ-5D utility score of 0.379. This is lower than the average score reported in HIV infected inpatients previously at QECH (0.498) [129], and substantially lower than the average score amongst HIV infected outpatients in the QECH catchment area (0.800) [115]. In my study, utility scores were an average of 0.125 lower in 3GC-R than 3GC-S infection. Mean VAS scores were 58.8 overall, with no substantial difference in scores between resistant versus sensitive infections (58.2 vs 59.5). Though easy to use, VAS scores (self-reported measure of symptoms recorded with a handwritten mark placed at one point along the length of a line) are subjective and not used by health economists to estimate quality-adjusted life years (QALYs) nor in economic evaluations.

I carried out multivariable analysis to determine the independent effect of 3GC-R on costs. In models adjusted for age, sex and causative organism, 3GC-R was associated with higher direct and societal costs as well as lower EQ-5D scores. I then included HIV as a model covariate, because previous

work from QECH has shown that HIV infected individuals have significantly higher health provider costs and significantly lower quality of life scores than HIV uninfected patients [129]. The estimated mean differences in HIV adjusted and HIV unadjusted models were very similar, (though with wider credible intervals), suggesting that in this cohort, HIV has only a marginal effect on costs and that 3GC-R is likely to be of greater significance.

I used annual incidence data for the two most commonly isolated Gram negative organisms in this cohort, *E. coli* and *Klebsiella* spp.[3], to estimate the economic burden of these bloodstream infections in Malawi. The estimates generated suggest an annual health provider spend of US$1,521, 628 on just these two infections in 2019. Over 80% of this spend is accounted for by 3GC-R, therefore strategies which aim to reduce AMR are likely to have profound impacts on healthcare spending.

To my knowledge, this study provides the first estimates of the QALYs lost due to AMR in the African region. QALYs provide a summary estimate of the quality and quantity of years lived with a disease. Estimating the QALYs lost allows us to understand the burden posed by that disease and is therefore an important metric for policy makers and academics to inform priorities for public health action and research. Globally, policy makers often use disability-adjusted life years (DALYs) for this purpose. However, both QALYS and DALYs are comparable metrics (1 DALY is equivalent to 1 QALY)[136] and are often used interchangeably. Although there are no estimates for the impact of AMR on DALYs or QALYs lost for the African region, estimates have been

made for the European region[140]. These European estimates suggested that AMR bacterial infections accounted for 170 DALYs per 100,000 population and that 3GC-R *E. coli* and *Klebsiella* accounted for over half of this burden. This is more than the combined burden estimates for three major infectious diseases (influenza, TB and HIV) in Europe [228]. Using the data generated in my study, I found a substantially higher burden, with approximately 200 QALYs are lost per 100,000 Malawians, from each of 3GC-R *E. coli* and 3GC-R *Klebsiella* alone. The high burden of AMR infections seen in Malawian neonates and children[3] and the associated high mortality, may in part explain this difference.

5.5.1 Limitations

This study has a number of limitations. Primarily, the assumptions made in the calculation of economic burden, will have generated conservative estimates, for the following reasons. Firstly, I have assumed that levels of 3GC-R remain at 2016 levels, when current trends suggest they will increase annually[3]. Secondly, in calculating the overall healthcare provider burden, I used costs estimated amongst adult admissions, and applied this to incidence data for patients of all ages. Paediatric admissions at QECH are likely to incur higher costs than adult admissions, because of the availability of more complex interventions (such as mechanical ventilation and central venous access). Primary costing studies on paediatric wards should therefore be a focus for future work, particularly since children carry a large proportion of the morbidity and mortality burden of 3GC-R (Chapter 4). Finally, burden estimates have been derived from cost data collected from one hospital only. Expanding economic

studies to other hospitals and healthcare settings within Malawi will help to generate more precise estimates.

5.6 Conclusions and future work

In this chapter, I have shown that 3GC-R BSI incur higher costs to the hospital and patients and lead to poorer HRQoL outcomes than 3GC-S infections. To my knowledge, I have generated the the first AMR economic burden estimates for sub-Saharan Africa, to be based on accurate, prospectively collected costing data. Strategies that reduce the incidence of 3GC-R infections could lead to significant cost savings to the hospital and patients, as well as improved QoL outcomes in those admitted to hospital. Future work should aim to improve the estimates I have generated, by incorporating costing data from paediatric wards and from wider healthcare settings.

5.7 Appendix

Appendix Table 5.1: Historical and projection population for Blantyre and Malawi.

Year	Blantyre Population (all ages)	Malawi Population (all ages)
1998	499,000	9,933,868
1999	514,000	10,152,753
2000	529,000	10,475,257
2001	544,000	10,816,294
2002	559,000	11,174,648
2003	575,000	11,548,841
2004	592,000	11,937,934
2005	609,000	12,341,170
2006	626,000	12,757,883
2007	644,000	13,187,632
2008	663,000	13,077,160
2009	682,000	13,512,376
2010	701,000	13,947,592
2011	721,000	14,388,550
2012	742,000	14,844,822
2013	763,000	15,316,860
2014	785,000	15,813,646
2015	808,000	16,310,431
2016	831,000	16,859,977
2017	855,000	17,409,522
2018	879,000	17,959,068
2019	905,000	18,508,613
2020	932,000	19,104,275
2021	962,000	19,718,415
2022	995,000	20,350,670
2023	1,031,000	20,350,670
2024	1,071,000	21,669,048
2025	1,114,000	22,358,192
2026	1,161,000	23,067,018
2027	1,213,000	23,794,786
2028	1,268,000	24,540,844
2029	1,326,000	25,305,919
2030	1,389,000	26,090,975

Appendix Table 5.2: Observed and estimated sensitive and resistant infections

	Blantyre		Malawi		Blantyre		Malawi	
Year	*E. coli* resistant	*E. coli* grown	*E. coli* resistant	*E. coli* grown	*Klebsiella* resistant	*Klebsiella* grown	*Klebsiella* resistant	*Klebsiella* grown
1998	115	185	2289	3683	66	146	1,314	2,907
1999	118	150	2331	2963	31	113	612	2,232
2000	83	116	1644	2297	38	101	752	2,000
2001	96	124	1909	2465	17	73	338	1,451
2002	113	147	2259	2939	20	81	400	1,619
2003	110	152	2209	3053	24	52	482	1,044
2004	92	133	1855	2682	21	46	423	928
2005	121	181	2452	3668	22	46	446	932
2006	84	186	1712	3791	18	52	367	1,060
2007	81	136	1659	2785	27	57	553	1,167
2008	81	116	1598	2288	36	55	710	1,085
2009	60	106	1189	2100	29	40	575	793
2010	64	96	1273	1910	34	45	676	895
2011	76	106	1517	2115	35	48	698	958
2012	42	86	840	1721	32	38	640	760
2013	74	100	1486	2007	33	55	662	1,104
2014	63	105	1269	2115	41	51	826	1,027
2015	65	118	1312	2382	39	48	787	969
2016	92	133	1867	2698	77	84	1,562	1,704
2017			1927	2786			1,613	1,760
2018			1988	2874			1,664	1,815
2019			2049	2962			1,715	1,871
2020			2115	3058			1,770	1,931
2021			2183	3156			1,827	1,993
2022			2253	3257			1,886	2,057
2023			2253	3257			1,886	2,057
2024			2399	3468			2,008	2,190
2025			2475	3578			2,072	2,260
2026			2554	3692			2,137	2,332
2027			2634	3808			2,205	2,405
2028			2717	3928			2,274	2,481
2029			2802	4050			2,345	2,558
2030			2889	4176			2,418	2,637

Chapter 6

Carriage of 3GC-R *E. coli* in Malawian households

6.1 Overview

In this chapter, I present the results of the longitudinal household cohort, which aimed to describe the drivers of 3GC-R carriage in a healthy population living in urban Blantyre. In total, 455 individuals in 110 households, sampled using spatially weighted randomisation, were recruited. Participants were followed-up at four time points, over 6 months. In total, 1,574 stool samples were collected and cultured for 3GC-R *E. coli*, using selective chromogenic agar. Overall, 3GC-R *E. coli* grew in 382/1574 (24.2%) samples.

Mixed-effects models were used to describe the associations of 3CG-R colonisation. The models used household and individual level random effects, in order to investigate the effect of household structure and individual level covariates on 3GC-R *E. coli* carriage status over time. In multivariable models, sample collection during rainy season and higher prevalence within households were associated with 3CG-R carriage. In univariable models a number of environmental associations, such as use of unprotected water sources and rubbish disposal in communal bins or rivers, were also important. A household member on CPT was associated with 3CG-R carriage on univariable analysis. These data suggest that community 3GC-R transmission is occurring and that environmental and

within-household drivers are important. Dynamical transmission models are needed to help elucidate the relative importance of within household person-to-person transmission from exogenous acquisition from a shared environmental source.

6.2 Introduction and chapter aims

The preceding three chapters have focussed on the risk factors and outcomes associated with invasive infection with 3GC-R organisms, in a cohort of hospitalised patients, demonstrating adverse clinical and economic outcomes. Gut mucosal colonisation with 3GC-R organisms is a risk factor for invasive infection and in this chapter I describe the analysis of my household cohort, which was designed to investigate the associations of 3GC-R *E. coli* colonisation in heathy individuals living in urban Blantyre. The specific aim of this chapter is to describe the drivers of ESBL *E. coli* carriage in urban Blantyre using longitudinal microbiological surveillance of humans and investigation of household and individual level risk factors

6.3 Methods

The epidemiological and laboratory methods for the household cohort are described in Chapter 2, General methods. Here, I focus on a description of the statistical methods used for data analysis.

6.3.1 Statistical analysis

Household and individual characteristics are displayed in tables as medians and interquartile ranges (continuous variables) or as proportions with

exact binomial 95% confidence intervals (discrete variables)[170]. The variables were selected a priori, based on clinical judgement, as potentially associated with 3GC-R gut mucosal colonisation.

The proportion of individuals who were colonised with 3CG-R *E. coli* at each time point, is expressed as a table of simple proportions. To investigate the degree of variation of 3GC-R status within individuals, the proportion of people who changed their status (positive to negative or negative to positive) at each time point was calculated and is presented as a table of proportions. 3GC-R *E. coli* carriage was also visualised as a heat map, showing changing status of individuals clustered in households, over time.

To investigate associations of household and individual characteristics on 3GC-R *E. coli* colonisation, logistic regression models were then constructed. Simple linear regression assumes independent observations of individuals, but in this dataset, individuals are grouped within households and followed longitudinally, with repeated sampling and covariate measurement at each time point. There is likely to be correlation between repeated measurements on the same individuals or same households over time, potentially exaggerating modelled associations. I therefore used mixed-effects models which include random effect terms for individuals and households.

The use of the random effect terms, allows for modelling of "missing" or unmeasured information on individuals and households, modelled as random samples from a distribution, and thus induces correlation amongst the same individuals and same households over time [175].

To begin with, associations between 3GC susceptibility status and the proposed predictor variables were screened for with univariable mixed-effects regression models, using the glmer function in R. Each individual or household variable was considered alone as a predictor, with an individual random-effect term used to account for correlation between individuals in the same household over time. Model structure was as follows:

$y_{ij} \sim \text{Bernoulli}\ (p_{ij})$

with

$$\text{logit}(p_{ij}) = \alpha + \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{T}\boldsymbol{\gamma} + u_i$$

where y_{ij} takes the value of 1 if individual i in housheold j is 3GC − R positive, and 0 otherwise. $\boldsymbol{x}_{ij}$ is a vector of individual-level covariates with coefficient vector $\boldsymbol{\beta}$, $\boldsymbol{z}_j$ is a vector of household-level covariates with coefficient vector $\boldsymbol{\gamma}$, and $u_j \sim N(0, \tau^2)$ is a random effect for the ith individual.

Estimates of the regression coefficients are provided as odds ratios (OR) with 95% confidence intervals (95% CI) and p-values are presented in the relevant tables.

To adjust for confounding and to produce unbiased risk-factor estimates, multivariable logistic regression models were then constructed, using predictor variables for which $p<0.20$ on univariable analysis. Three models were constructed to arrive at the final, best fitting multivariable model. First, all predictor variables selected in univariable screening were included

in a generalised linear regression model without random effects (Model 1). Significant variables from this model were taken forward to the second model – a mixed effects model with two random effect terms: one for individuals and one for households (Model 2). A third model, replaced the household random effect with a linear continuous variable measuring household 3GC-R prevalence. This prevalence term was calculated as the number of further people in the household who were positive for 3GC-R and used as an alternative and more parsimonious measure of the effect of household clustering on 3GC-R status (Model 3). This latter multivariable model was the best performing model, using Akaike information criterion (AIC), as a measure of prediction error, and was therefore used to make the final estimates of association. Resulting odds ratios and p-values are tabulated.

6.4 Results

6.4.1 Participant characteristics

Baseline household visits were conducted between 11th April 2018 and 21st March 2019 and follow-up for all households was completed by 21st September 2019. 110 households were recruited at baseline. Figure 6.1 shows the 110 household sample points as locations within urban Blantyre. Figure 6.2 shows participant flow across the 6-month follow-up period, including number of samples collected and drop outs. A total of 455 samples from individuals in 110 households were collected at baseline. At day 30, samples were collected from 396 individuals in 103 households. At day 90, samples were collected from 374 individuals in 100 households and by day 180, this was 349 samples in 99 households. Seven full households

withdrew from the study due to household relocation and a further four households chose to withdraw from the study before follow-up was complete (Figure 6.2). 59/455 (13.0%) of samples were missing at day 30, 81/455 (17.9%) were missing by day 90 and 106/455 (23.3%) were missing by day 180.

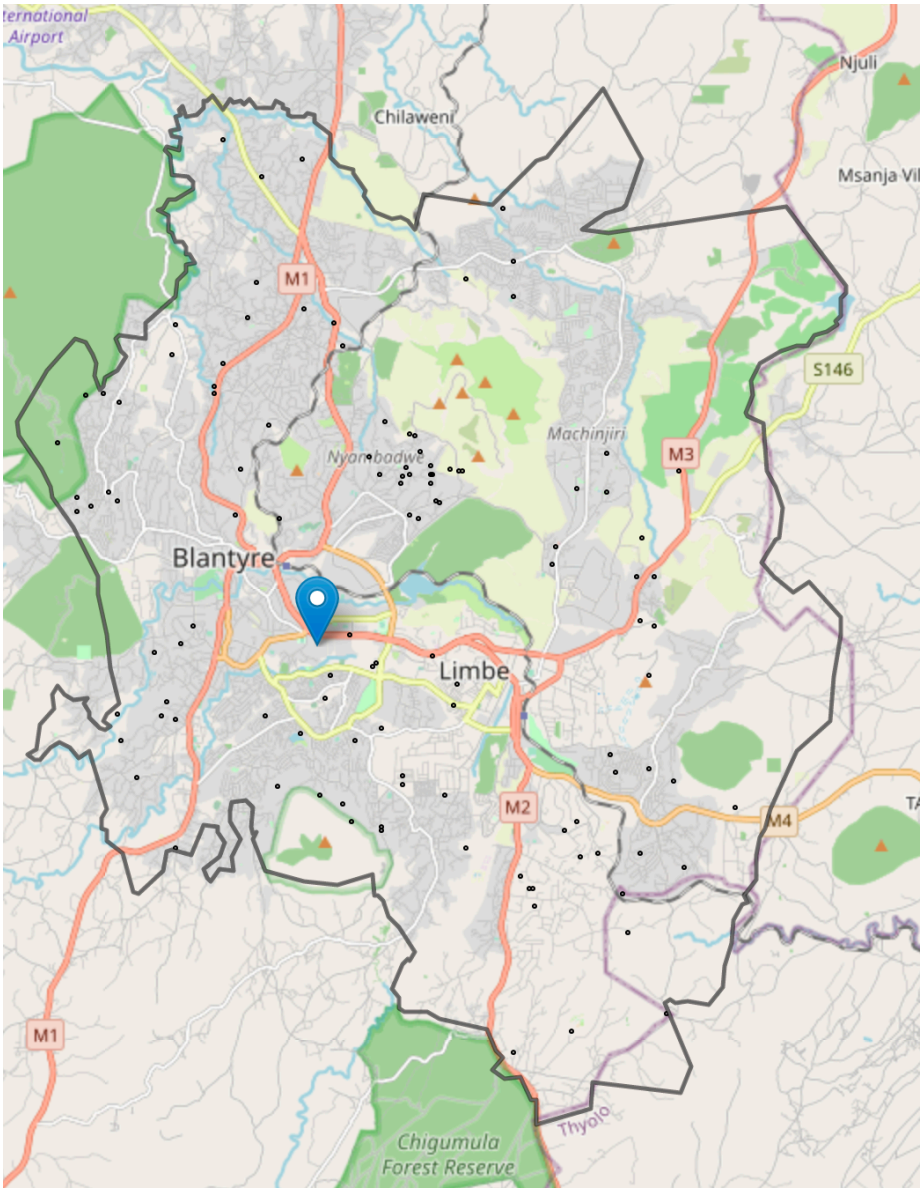


Figure 6.1: Sample points of the 110 households shown as black circles. Points are shown with jitter (addition of small amount of noise) for confidentiality. Blue marker shows location of QECH. Produced in R using openstreetmap.org.

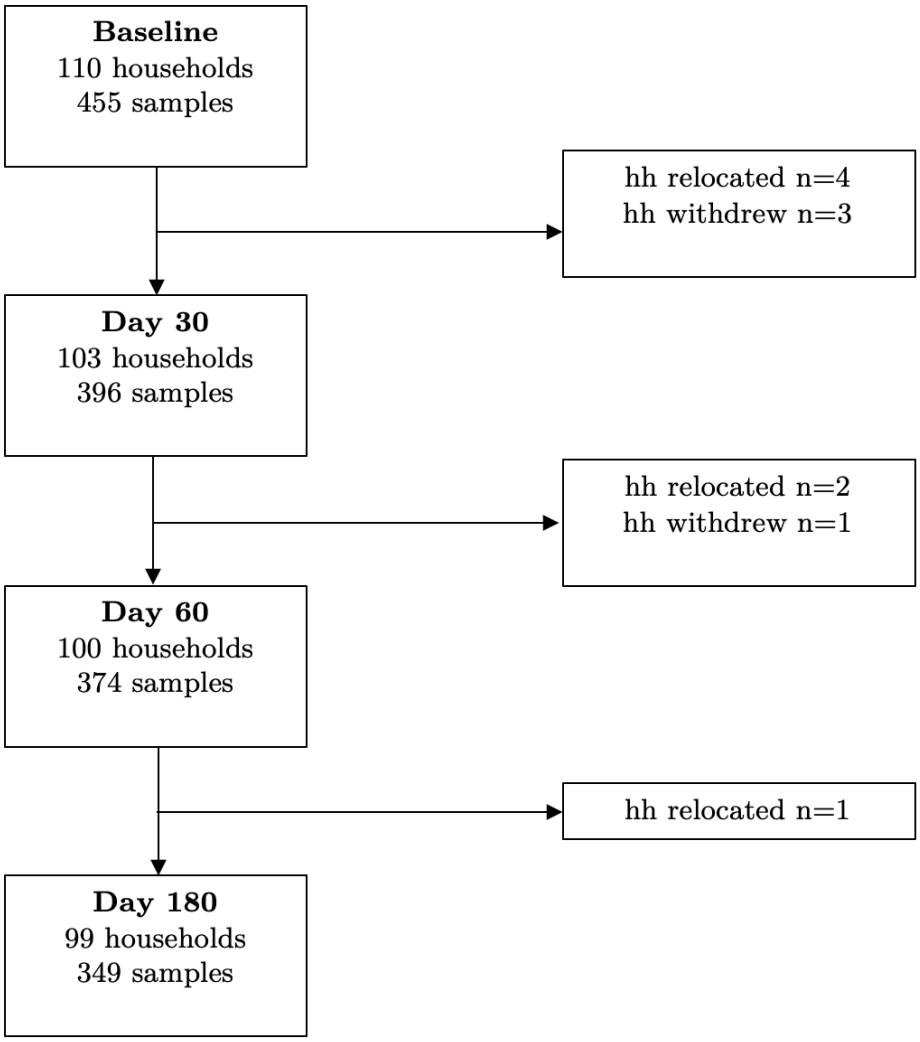


Figure 6.2: Study population at baseline and follow-up. hh=household
Note: 1 sample = 1 individual

Individual characteristics for participants enrolled at baseline, are shown in Table 6.1. Many more participants were female than male, (82.2% vs 18.2%), likely reflecting that more women were at home during the weekday working hours that the study team were recruiting. Participants had a median age of 15.3 years (6.8-28.2). At baseline, 247/455 (54.3%) were children under 16 and of these, 91/455 (20.0%) were under 5 years old. HIV status was known for 275/455 (60.1%) of participants and of

these, 37/275 (13.5%, 95%CI, 9.7-18.1) were HIV infected. Of those with HIV infection, 34/37 (91.9%, 95%CI, 78.1-98.3) were on ART.

Table 6.1: Individual characteristics

Variable	Total
Demographics	
Age, Median (IQR)	15.3y (6.8-28.2)
Female sex	374/455 (82.2[78.4-85.6])
Male sex	81/455 (17.8[14.4-21.6])
HIV status	
HIV positive	37/275 (13.5[9.7-18.1])
HIV negative	235/275 (85.5[80.7-89.4])
HIV unknown	180/455 (39.6[35.0-44.2])
HIV exposed/uninfected in <18m	3/23 (13.0[2.8-33.6])
Current ART (if HIV infected)	34/37 (91.9[78.1-98.3])
Current CPT (if HIV infected or exposed)	35/40 (87.5[73.2-95.8])
Antibiotic use	
Prior to baseline visit*	73/455 (16.0[12.8-19.7])
Baseline to day 30	24/396 (6.1[3.9-8.9])
Day 30 to day 90	30/374 (8.0[5.5-11.3])
Day 90 to day 180	47/349 (13.5[10.1-17.5])

Note:

* 3 months prior to study visit
IQR=Interquartile range, ART = Antiretroviral therapy, CPT = Cotrimoxazole preventive therapy.

In total, 73/455 (16.0%, 95%CI, 12.8-19.7) participants had received at least one course of antibiotics within the 3 months prior to the baseline visit. Between the baseline and day 30 visit, this was 24/396 (6.1%, 95%CI, 3.9-8.9) and between day 30 and 90, this was 30/374 (8.0%, 95%CI, 5.5-11.3). Between the day 90 and day 180 visit this was 74/349 (13.5%). Figure 6.3 shows the antibiotics used by study participants during the study period. Co-trimoxazole (not including CPT) and amoxicillin accounted for 77.7% of all antibiotic courses used.

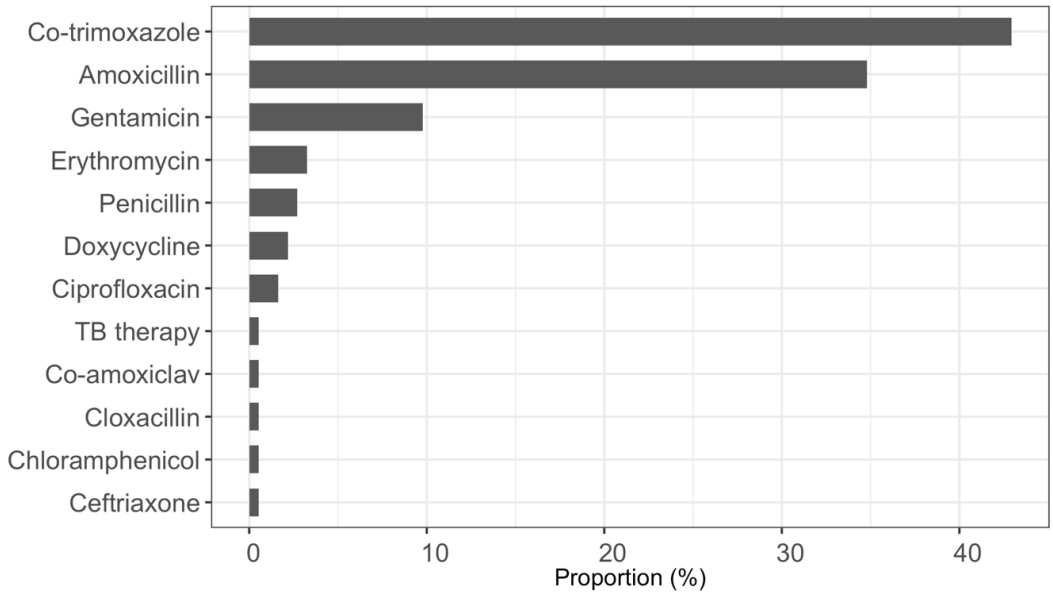


Figure 6.3: Antibiotics used by individuals during the study period. Co-trimoxazole does not include CPT

WASH and healthcare exposure characteristics of the 110 households enrolled at baseline are shown in Table 6.2. Over 60% (69/110 [62.7%, 95%CI, 53.0-71.8]) of households shared their main toilet with at least one other household (median of 3 households) but there was little heterogeneity in terms of toilet type, with 96/110 (87.3%, 95%CI, 79.6-92.9) of households using an ordinary pit latrine. 17/110 (15.5%, 95%CI 9.3-23.6) of households had access to tap water inside their homes, with the reminder accessing water from sources outside their homes including communal boreholes or wells (Table 6.1). On direct questioning, most participants said they were able to wash their hands after toilet use, with just over half (61/110 [55.5%, 95% CI, 45.7-64.9]) saying they used soap. Fewer than one-quarter of households had soap 27/110 (24.5%) or water 23/110 (20.9%) available at the place of handwashing when observed by the study team.

At baseline, 22/110 (22%, 95%CI, 13.0-28.7) of households had at least one household member on CPT. At the baseline visit, 11/110 (10.0%) of households reported at least one member with current diarrhoea and 12/110 (10.9%) reported at least one member with current fever. A higher proportion of participants reported that at least one member had diarrhoea or fever the last 3 months (diarrhoea. - 30/110 [27.3%]; fever 44/110 [40.0%].

Table 6.2: Household characteristics

	Total
No. in household	5(4-6)
No. rooms in house	4(3-5)
Animals at home	
Dogs	16/110 (14.5 [8.5-22.5])
Poultry	18/110 (16.4 [10.0-24.6])
Livestock	2/110 (1.8 [0.2-6.4])
Other animals	4/110 (3.6 [1.0-9.0])
No animal	70/110 (63.6 [53.9-72.6])
Any animal	40/110 (36.4 [27.4-46.1])
WASH	
Toilet type	
Flushing toilet	11/110 (10.0 [5.1-17.2])
VIP	1/110 1/110 (0.9 [0.02-5.0])
Ordinary pit latrine	96/110 (87.3 [79.6-92.9])
Composting	2/110 (1.8 [0.2-6.4])
Shared toilet†	69/110 (62.7 [53.0-71.8])
Households shared with†	3(2-5)
Open defaecation	16/110 (14.5 [8.5-22.5])
Water for drinking	
Piped into dwelling	17/110 (15.5 [9.3-23.6])
Piped outside dwelling	18/110 (16.4 [10.0-2.5])
Communal	65/110 (59.0 [49.3-63.8])
tap/standpipe	
Borehole	9/110 (8.2 [3.8-15.0])
Unprotected well/spring	1/110 (0.9 [0.02-5.0])
River	0
Treated drinking water	11/110 (10.0 [5.1-17.2])
Distance to water	
On premises	28/110 (25.5 [17.6-34.6])
<30min	75/110 (68.2 [58.6-76.7])
>30min	7/110 (6.4 [2.3-12.7])
Place of rubbish disposal	
Collected	10/110 (9.1 [4.5-16.1])
Communal bin	24/110 (21.8 [14.5-30.7])
Left in open	68/110 (61.8 [52.0-70.9])
Nearest river	9/110 (8.2 [3.8-15.0])
Hand washing	
Water available at place of handwashing	23/110 (20.9 [13.7-29.7])
Soap available at place of handwashing	27/110 (24.5 [30.8-49.8])
Participants state washing hands after toilet use	106/110 (96.4 [90.9-99.1])
Participants state using soap for hand washing after toilet use	61/110 (55.5 [45.7-64.9])
Healthcare and illness	
Household member on CPT	22/110 (20.0 [13.0-28.7])
Household member with current diarrhoea	11/110 (10.0 [5.1-17.2])
Household member with current fever	12/110 (10.9 [5.8-18.3])
Household member with diarrhoea in prior 3m	30/110 (27.3 [19.2-36.6])
Household member with fever in prior 3m	44/110 (40.0 [30.8-49.8])
Household wealth	
Wealth indicator score	1.99 (1.71-2.56)
Travel outside Malawi in prior 3 months*	12/455 (2.7[1.4-4.6])

Note: Numeric values are shown as proportions with exact binomial 95% CI or as medians with IQR. *All travel was to South Africa. WASH = Water, sanitation and hygiene, VIP = Ventilated improved pit latrine , CPT = Cotrimoxazole preventive therapy

6.4.2 3GC-R *E. coli* colonisation

Table 6.3 shows 3CG-R *E. coli* colonisation as a function of time. Prevalence was similar at baseline (21.1%, 95%CI, 17.4-25.1), day 28 (21.7%, 95% CI 19.1-26.1) and day 90, (23.3%, 95% CI 19.1-27.9), but was higher by the final day 180 visit (32.4%, 95%CI, [27.5-37.5]).

Table 6.3: 3GC-R *E. coli* carriage at each time point

Visit	n	Prevalence 3GC-R n/N (%)
Baseline	455	96/455 (21.1[17.4-25.1])
Month 1	396	86/396 (21.7)[19.1-26.1])
Month 3	374	87/374 (23.3)[19.1-27.9])
Month 6	349	113/349 (32.4)[27.5-37.5])

Within individuals over time, there were frequent status changes between colonised and un-colonised states (Table 6.4 and Figure 6.4). Almost one-third of individuals had a status change between the 28 and 90 day visits (28.3%, 95%CI, 23.9-33.0) and just over a third had a status change between day 90 and day 180 (33.5%, 95%CI, 28.6-38.7). Of the 349 individuals who gave samples at all time points, 118/349 (33.2%, 95% CI, 28.9-39.0) kept the same 3GC-R status throughout. Of these, 112/118 (95.0%, 95%CI, 89.3-98.1) were always negative and 6/118 (5.0%, 95%CI, 1.9-10.7) were always positive. Two households remained negative at all 4 time points, but there were no distinguishing features of these households.

Table 6.4: 3GC-R *E. coli* status change between visits

Visit	n	3GC-R n/N (%)
Baseline – Say 28	396	112/396 (28.3[23.9-33.0)
Day 28 – Day 90	374	104/374 (27.8[23.3-32.6])
Day 90 – Day 180	349	117/349 (33.5 [28.6-38.7])

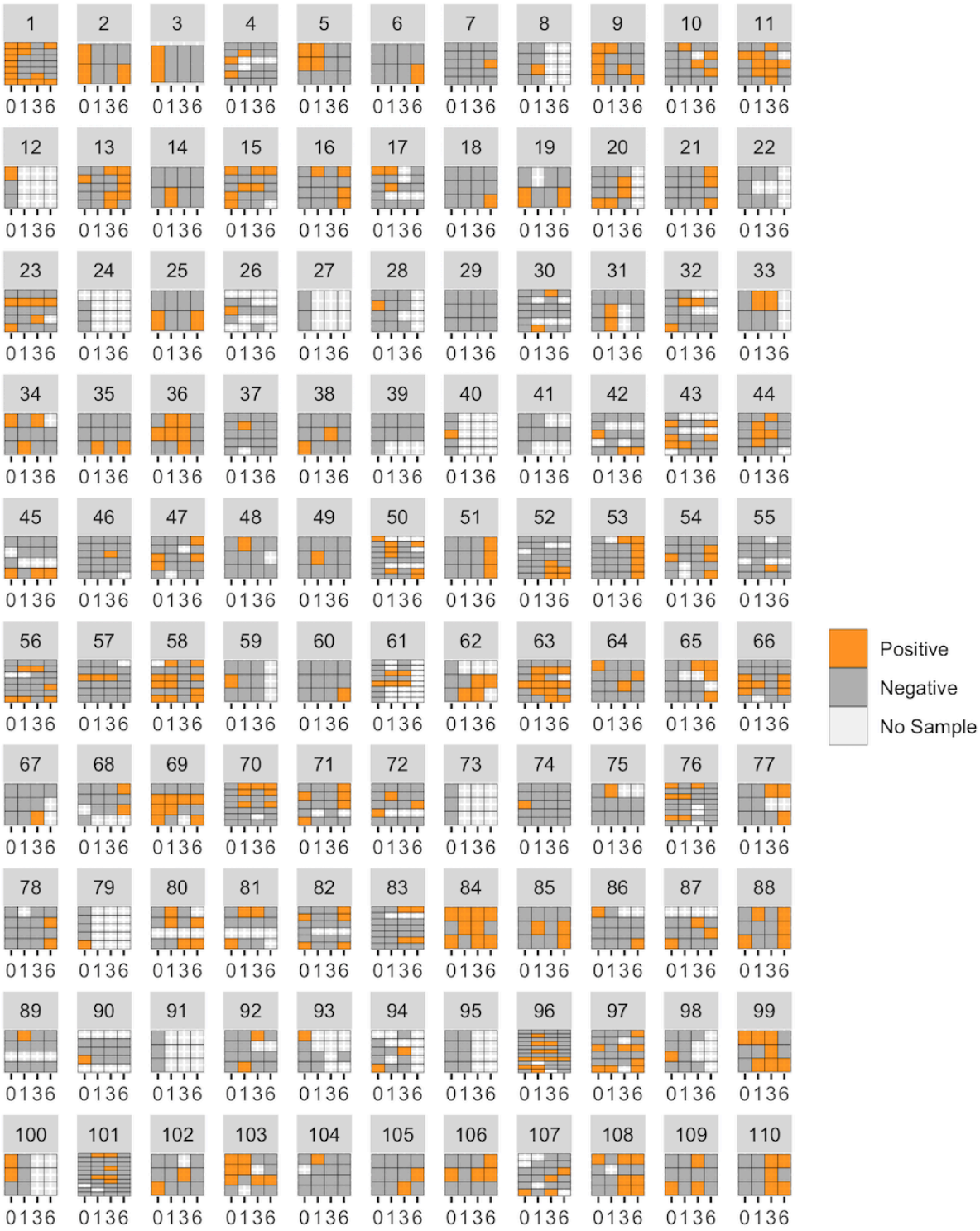


Figure 6.4: Facet plot showing the distribution of stool samples which were positive for 3GC-R *E. coli* within individuals and households. Each cell represents a stool sample, coloured by presence of absence of 3GC-R *E. coli* (or missing sample). Each row represents an individual Each column represents a study visit. e (baseline=0, month1=1,

month3=3, month6=6). Each facet represents a household (numbered 1-110). The high degree of status switching is shown.

6.4.3 Associations of 3GC-R *E. coli* colonisation

Table 6.5 shows univariable associations of individual level variables on 3GC-R colonisation and Table 6.6 shows univariable associations of household level variables. Variables for which $p<0.2$, which were taken forward for analysis in the multivariable model, are shown in bold. On univariable analysis, sample collection during rainy season was associated with 3GC-R colonisation, OR 2.60(95%CI, 1.94-3.48), as was the use of unprotected water sources, OR 3.46 (95%CI, 1.15- 10.36). Prevalence of 3GC-R in the rest of their household was strongly associated with 3CG-R colonisation of an individual, (OR 26.6, 95%CI 12.8-55.2). Use of communal taps/standpipes as the main water source was negatively associated with 3GC-R, OR 0.69 (95%CI, 0.51-0.93), as was participants reporting use of soap for handwashing, OR 0.74 (95%CI, 0.55-0.99).

Other variables which had potentially significant effect sizes on univariable analysis, but with CIs crossing 1, included presence of a household member on CPT OR 1.39 (95%CI, 0.98-1.97) and presence of a household member with current fever, OR 1.48 (95%CI, 0.98-2.23). There was no interaction between current fever and recent antibiotic use on univariable modelling, suggesting that antibiotic use was not contributing to this association. Recent antibiotic use was not associated with 3GC-R colonisation on univariable analysis OR, 95%CI, 1.20(0.81- 1.79). However, because of the importance of antibiotic use as a risk factor for 3GC-R in other contexts [42], this variable was taken forward to multivariable analysis.

Place of rubbish disposal was potentially important. Having rubbish collected from the home showed no evidence of association with 3GC-R, but use of a communal bin (OR 1.32, 95%CI, 0.94-1.85) or the nearest river (OR 1.39 (95%CI, 0.85-2.27) were selected at the 20% Type 1 error rate. There was some evidence that having a household member who had travelled outside of Malawi (to South Africa) in the preceding 3 months, was associated with colonisation, though with wide confidence intervals, reflecting the small sample OR 2.42 (0.72-7.84).

Table 6.5: Univariable analysis of individual level variables

	3GC-R	p
Age group*	1.00 (0.99-1.01)	0.383
Sex	1.19 (0.82-1.71)	0.368
HIV or HIV exposed	0.99 (0.98-1.00)	0.074
ART	0.99 (0.99-1.00)	0.082
CPT	0.64 (0.26-1.55)	0.319
Recent antibiotic use**	1.20(0.81- 1.79)	0.038

Note:

* age showed non-linear association with 3CG-R so was categorised

** recent = 3 months preceding baseline visit, or between visits

Table 6.6: Univariable analysis of household level variables

No. in household	1.02(0.95-1.10)	0.467
No. rooms in house	1.00 (0.99-1.02)	0.618
WASH		
Toilet at home	1.01 (0.87-1.16)	0.940
Toilet type		
Flushing toilet	0.99 (0.61-1.63)	0.967
VIP	0.67 (0.15-3.27)	0.672
Ordinary pit latrine	1.08 (0.69-1.69)	0.710
Composting	0.21-2.29)	0.557
Shared toilet†	1.02 (0.76-1.38)	0.873
Open defaecation	0.99 (0.93-1.07)	0.940
Water for drinking		
Piped into dwelling	**1.31 (0.89-1.93)**	**0.163**
Piped outside dwelling	1.22 (0.80-1.87)	0.354
Communal tap/standpipe	**0.69 (0.51-0.93)**	**0.013**
Borehole/tube well	1.08 (0.65-1.81)	0.745
Unprotected well/spring	**3.46 (1.15-10.36)**	**0.027**
Place of rubbish disposal		
Collected	0.76 (0.45-1.33)	0.345
Communal bin	**1.32 (0.94-1.85)**	**0.107**
Left in open	0.97 (0.58-1.04)	0.860
Nearest river	**1.39 (0.85-2.27)**	**0.180**
Hand washing		
Water available at place of handwashing	1.38 (0.99-1.94)	0.540
Soap available at place of handwashing	1.16 (0.81-1.65)	0.419
Participants state washing hands after toilet use	1.46 (0.62- 3.45)	0.385
Participants state using soap for hand washing after toilet use	**0.74 (0.55-0.99)**	**0.042**
Sample collection in rainy season	**2.60(1.94-3.48)**	**<0.001**
Household member goes to school	0.98 (0.92-1.05)	0.598
Household member has fever currently	**1.48 (0.98-2.23)**	**0.065**
Household member has diarrhoea currently	1.02 (0.62-1.68)	0.949
Household member with fever in prior 3 months	0.98 (0.93-1.03)	0.448
Household member with diarrhoea in prior 3 months	1.00 (0.95-1.06)	0.910
Household member on CPT	**1.39 (0.98-1.97)**	**0.068**
Healthcare contact in prior 3 months	1.00 (0.57 -1.44)	0.686
Household member travelled outside Malawi in prior 3 months	**2.42 (0.72-7.84)**	**0.140**

Note:

Numeric variables are shown as proportions with exact binomial 95%CI or as medians with IQR. Variables for which p<0.2 on univariable analysis are shown in bold

WASH = Water, sanitation and hygiene. VIP = Ventilated improved pit latrine, CPT = Cotrimoxazole preventive therapy.

The results of multivariable analysis are shown in Table 6.7. In the final multivariable model (Model 3), only sample collection during rainy season and higher prevalence of 3GC-R in the rest of the household remained associated with 3GC-R *E. coli*. Other variables - household member on CPT and rubbish disposal in a river or communal bin were not strongly associated with 3GC-R colonisation, but confidence intervals contained a clinically relevant effect size.

Table 6.7: Multivariable associations of 3GC-R colonisation

	Model 1		Model 2		Model 3	
Variable		p		p		p
HIV infected or exposed	0.99 (0.99-1.00)	0.252	-	-	-	-
Household member with current fever	1.20 (0.60-2.33)	0.588	-	-	-	-
Water piped into dwelling	1.28 (0.74-2.21)	0.355	-	-	-	-
Water from unprotected well/spring	1.24 (0.32-4.79)	0.748	-	-	-	-
Soap used for handwashing	0.91 (0.61-1.3)	0.671	-	-	-	-
Household member on CPT	1.66 (1.02-2.71)	0.036	1.34 (0.87-2.06)	0.183	1.35 (0.88-2.08)	0.170
Sample collection in rainy season	2.26 (1.43-3.57)	<0.001	2.21 (1.55-3.15)	<0.001	2.20 (1.55-3.17)	<0.001
Rubbish disposal in river	2.61 (1.08-6.42)	0.003	1.55 (0.86-2.79)	0.149	1.56 (1.55-3.15)	0.144
Rubbish disposal in communal bin	2.62 (1.30-5.41)	0.008	1.46 (0.95-2.23)	0.078	1.45 (0.95-2.21)	0.085
Household prevalence	-	-	-	-	16.2 (7.65-34.2)	<0.001
Recent antibiotic use	1.00 (0.99-1.02)	0.567	0.99 (0.98-1.01)	0.623	1.11 (0.47-2.68)	0.805
Recent travel outside Malawi	1.86 (0.38-7.23)	0.389	-	-	-	-

Note:

Model 1 = generalised linear model

Model 2 = mixed effects model with household and individual random effects

Model 3 = mixed effects model where household prevalence replaces household random effect term

“Water piped into dwelling” removed from models due to collinearity with “water from well” variable

6.5 Discussion

In this chapter I have analysed data from a longitudinal cohort of households in urban Blantyre, to investigate the drivers of 3GC-R *E. coli* gut mucosal colonisation in Malawian adults and children. The data show that the prevalence of 3GC-R *E. coli* carriage is high and that environmental and household level risk factors are likely to be important drivers of transmission.

Community 3GC-R *E. coli* carriage prevalence ranged between 21% and 32% in this study, considerably higher than estimates from European settings, where prevalence is typically less than 10%[47-49]. As discussed in Chapter 1, there are few other community studies of 3GC-R colonisation from sSA, but the estimates I have generated are comparable to those from the systematic review of studies in sSA (pooled carriage prevalence 18%, 95%CI, 12-28)[29]. A recent longitudinal study of hospitalised patients in Blantyre, found a similarly high prevalence of 3GC-R Enterobacterales carriage amongst community controls (median 28%) (Lewis, Submitted).

The associations of 3GC-R described in this chapter are supportive of community and household level transmission. A higher prevalence of carriage within the same house, was strongly associated with 3GC-R colonisation, suggesting that within-household transmission may be occurring. WASH factors, such as use of unprotected water sources and open rubbish disposal were associated with 3CG-R in univariable and multivariable modelling, suggesting an important role for environmental risk-factors. The strong seasonal effect on carriage (rainy season association

with 3GC-R), is also consistent with an environmental route of transmission. Blantyre is a densely populated, urban setting, often with poorly maintained sewerage systems and a reliance on pit latrines and unprotected drinking water sources. Heavy rains are known to increase faecal contamination of water in these environments and seem therefore to be contributing to increased 3GC-R carriage rates[229]. As described in Chapter 1, there are few other longitudinal cohorts from sSA to provide insight into temporal carriage trends and all are healthcare facility based. However a recent hospital based longitudinal carriage study which recruited community controls in Blantyre, found a similar rainy season association (Lewis, in press).

It is interesting that antibiotic use, though a relatively frequent exposure (16% of people had an antibiotic course in the 3-months prior to the baseline visit), was not associated with 3GC-R colonisation. In other studies from sSA, antimicrobial usage has been identified as a risk factor, though this is usually in the context of recent hospitalisation, which may confound this antibiotic effect[29]. CPT, however, may play a more important role. A household member on CPT was associated with 3GC-R on univariable analysis, though being HIV infected was not. It is difficult to draw firm conclusions from this lack of HIV association, since HIV status was unknown in 40% of the cohort. As discussed in Chapter 1, there are few other studies which investigate the effects of HIV and CPT on 3GC-R carriage (one study of Tanzanian children found HIV to be significantly associated with 3GC-R)[41]. Perhaps it is the combination of the direct effect of HIV on gut mucosal function, combined with the effect of CPT on gut microbiota that is important, rather than HIV or

antimicrobials alone. It may also be that the consistent exposure from a daily, life-long CPT dose places more selection pressure for 3GC-R *E. coli* carriage, than does short term antimicrobial courses. Given the widespread usage of CPT in Malawi[152] and its significant mortality benefit on HIV infected individuals[230], this association warrants further investigation, and if a consistent finding in other studies, further consideration in terms of risk-benefit analysis.

Prior hospitalisation has been identified as a risk factor for 3GC-R carriage in other studies from sSA[231-234], but recent hospitalisation (either as a patient or guardian) was a relatively rare exposure in this study (7.9% at baseline). Facility based studies are generally required to examine the impact of healthcare on colonisation[42].

6.5.1 Limitations

There are a number of limitations to this study and analysis. The study population was 80% female, reflecting that more women than men were at home during the week day working hours that samples were collected and perhaps reluctance of men to participate. The study team attempted to overcome this by leaving sample pots with families overnight, but this did not improve male participation in the study. Out of hours recruitment may help to solve this problem, but first the reasons for lack of male participation need to be investigated further. Given the strong effect of household prevalence on carriage however, it is unlikely that this has had significant impact on the overall results.

The proportion of missing samples was fairly high and many were due to transfer out of the study area. Given that this mobile population is of interest in terms of 3GC-R exposures, this is a potential source of bias in the study.

In general, the study team relied on self-reported exposures, including HIV, antimicrobial use and hospitalisation, potentially leading to bias. An attempt was made to verify responses in 'health passports' which are kept at home and contain an individual's health record, however these were only available for 33% of participants. There are a number of other risk-factors which may be important in 3GC-R acquisition, but which were rare exposure in this cohort. Travel, for example, showed an association with 3CG-R in univariable analysis, but with wide confidence intervals reflecting a small sample size.

Detection of 3CG-R *E. coli* in this study relied on selective culture with chromogenic agar. Whilst this is a sensitive and specific technique, it may not detect all 3GC-R producers[235]. The highly complex and dynamic switching between 3GC-R positive and negative states may be real, or it may be a function of inadequate test sensitivity. Additionally, the time periods between sampling were chosen for largely pragmatic reasons, but gaps in sampling may have missed acquisitions and losses.

Although I have used random effects to account for clustering of households and individuals over time, regression models are far from ideal at analysing longitudinal data. The assumptions made in traditional regression, such as absence of independence between individuals and a

changing prevalence of infected and susceptible hosts over time, do not necessarily hold true for 'infections' (or colonisation events). Whilst the data presented here are supportive of community and household 3GC-R drivers of colonisation, it is not possible to elucidate the relative importance of within household person-to-person transmission from exogenous acquisition from a shared environmental source. For example, are individuals acquiring 3GC-R directly from a person in their household who is on CPT, or is the whole household becoming colonised from use of a contaminated water source? Dynamical transmission models can overcome some of the limitations of regression and I will attempt to address some of these issues in the following chapter, by fitting a transmission model to the household cohort data.

6.6 Conclusions and next steps

Gut mucosal colonisation with 3GCR-E is a risk factor for subsequent infection[26, 27] and healthy colonised individuals are therefore thought to be a reservoir for spread of these organisms, contributing to onward transmission and disease[28]. Prevalence of 3GC-R colonisation in Malawian adults and children was high and interventions aimed at interrupting 3GC R transmission are needed.

I show that within household transmission is important and acquisition from environmental routes is likely. The relative importance of person to person transmission and exogenous acquisition from environmental sources cannot be determined from the current analysis and will be explored using a dynamical transmission model in the following chapter.

Chapter 7

Dynamical transmission models of 3GC-R *E. coli* carriage

7.1 Chapter overview

In this chapter I develop the framework for a dynamical transmission model of 3GC-R *E. coli* carriage and fit the model to data derived from the household cohort described in Chapter 6. I use a stochastic, discrete time model, which parameterises the household network, with the aim of exploring the relative importance of within household person-to-person transmission versus shared transmission from an environmental source.

I hypothesise two different model structures and use a chain binomial algorithm to simulate the epidemic process. I manually explore parameter values and visually fit the simulated data to observed household 3GC-R *E. coli* prevalence, making comparisons between the two model outputs. I show that the within household network may be a key driver of transmission against which interventions should be targeted and subsequently describe how the models should be refined and progressed in future work.

7.2 Introduction and chapter aim

In Chapter 6, I showed that 3GC-R colonisation amongst adults and children in urban Blantyre, is a highly dynamical process, with individuals switching status multiple times over the study period. I used hierarchical models to demonstrate that community level, household transmission is important and describe a strong seasonal effect, suggestive of acquisition from environmental routes.

The longitudinal models used in Chapter 6, however, make a number of assumptions which render them inadequate for describing the process of 3GC-R colonisation, including assumptions of independence between individuals and of a fixed prevalence of colonised and uncolonised individuals over time[175, 236]. In order to understand where best to target interventions aimed at reducing 3CG-R colonisation, it is necessary to investigate this household transmission in more detail, by using models which are able to represent the dynamical process of colonisation more accurately.

As I discuss in Chapter 1, there are few community-based studies which use dynamical modelling of transmission in this way. Briefly, one study from the Netherlands, used a transmission model fit to hospital and household data, showing evidence of within household transmission[69].The second was a population-based transmission model which fit 3GCR *E. coli* data from multiple sources (animals, food, humans, water), finding evidence of person-to-person transmission had a greater impact on

transmission than direct animal contact, food consumption or environmental sources[70].

Based on the results in Chapter 6, there are two main hypotheses for where transmission is occurring:

1. Transmission is predominantly occurring from environment to person and vice-versa. An environmental reservoir will therefore persist unless specifically targeted;
2. Transmission is predominantly occurring from person-to-person, within a household. Interrupting individual level transmission within the house will eliminate the environmental reservoir.

This chapter aims to develop a transmission model which can ultimately be used to explore these potential transmission routes in more detail. The specific aims of this chapter are as follows:

To develop a plausible model of 3GC-R *E. coli* dynamics in the community, driven by the broad determinants of environment-to-person and person-to-person transmission;
To use this model to explore the relative importance of environmental versus person-to-person acquisition of gut mucosal 3GC-R *E. coli*.

7.3 Methods

7.3.1 Model framework

The first step in model construction is to identify the key biological processes involved in 3GC-R colonisation. In Chapter 6, I showed that within household transmission and rainy season are key risk factors for carriage and in what follows, I will abstract these processes into a non-linear mathematical equation, to measure force of infection, i.e. the per capita rate at which uncolonised individuals become colonised. I will then investigate the effects of changing transmission parameters in the equation, on the overall force of infection from uncolonised to colonised states, aiming to derive a set of parameters which best fit the observed data from the household cohort.

The model characterises individuals within a closed population as either uncolonised (U) or colonised (C), Figure 7.1. Individuals are assorted into households, within and between which we assume they mix homogeneously. At any time, t, uncolonised individuals receive a seasonally-varying force-of-infection, $\lambda_{j(t)}$, from colonised individuals parameterised in terms of a "global" force-of-infection, β_G, between all individuals in the population plus a "household" rate, β_{HH}, giving the extra force of infection experienced from colonised members of the household. Colonised individuals are assumed to lose their 3GC-R status, returning to the uncolonised state at rate, $\gamma_{j(t)}$. The model works in weekly time-steps assuming constant UC and CU rates respectively according to the chain-binomial setup[237] described below.

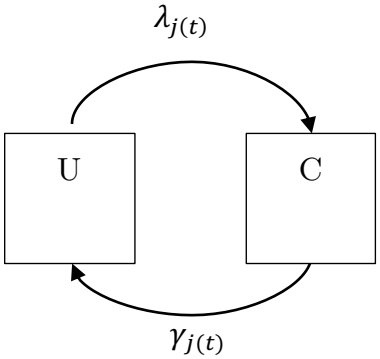


Figure 7.1: Two state model of 3GC *E. coli* acquisition, showing two transitions: acquisition, λ, and loss, γ for a set of individuals, j, who are in colonised (C) or uncolonised (U) state at time, t.

Two colonisation rate models are considered. The different models represent two plausible scenarios for the effect of season on 3GC-R transmission. In Model 1, we aim to examine the effects of season on transmission within and outside the house. It is possible, however, that the within household environment is relatively protected from seasonal exposures (rain, UV light etc) and so Model 2 represents a scenario where only external transmission is affected by season.

Model 1: seasonality is applied to household and global transmission parameters, such that:

$$\lambda_j(t) = 1 - cos(2\pi t/365)\left[\beta_G \frac{C(t)^T 1}{N} + \beta_{HH} \sum_{i \in C(t)} h_{ij} C_i(t)\right]$$

Model 2: seasonality is applied to global transmission parameter only, such that:

$$\lambda_j(t) = 1 - cos(2\pi t/365)\left[\beta_G \frac{C(t)^T 1}{N}\right] + \beta_{HH} \sum_{i \in C(t)} h_{ij} C_i(t)$$

Where:

$\lambda_j(t)$is the force of infection on uncolonised individual (j) at time (t)

The time period t: t=1,…,T where T=105

$1 - cos(2\,\pi t/365)$ is a seasonal parameter modelled using a harmonic term

β_G (global transmission parameter) and β_{HH} (household transmission paramter) are the parameters to be estimated

C(t) is a vector of individuals in the colonised state at time, t and $h_{ij}C_i(t)$ is a household mixing matrix with $h_{ij} = 1$ if individuals i and j share membership of a household and $h_{ij} = 0$, otherwise. The mixing matrix is shown in in Figure 7.2.

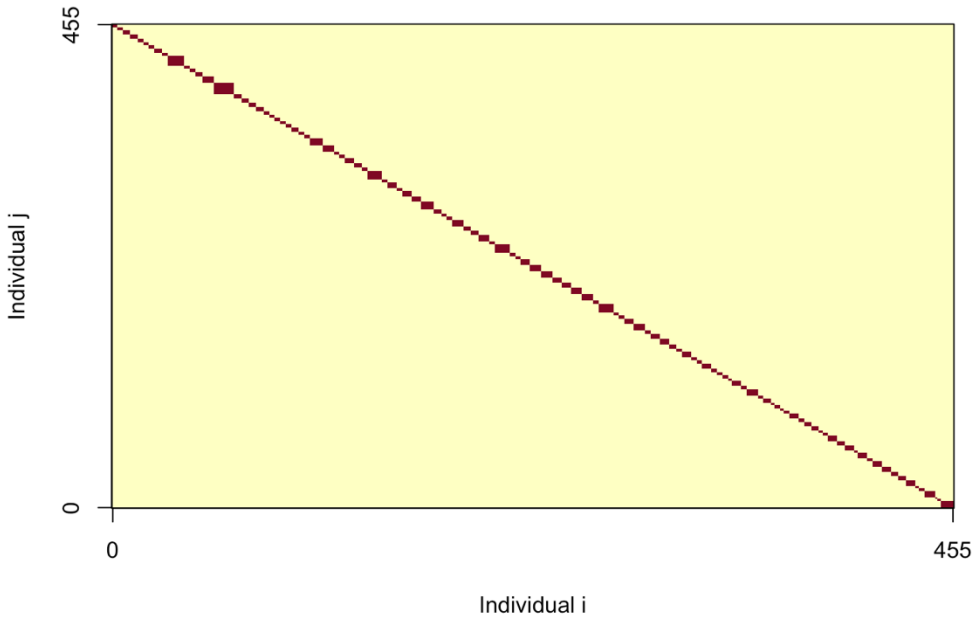


Figure 7.2: Graphical representation of the household mixing matrix. Red elements indicate individuals in the same household. Cream elements indicate individuals who are not in the same household. The contact pattern is therefore described by the mixing matrix which has 455 rows and columns, representing each individual recruited at baseline.

7.3.2 Model Simulation

A chain binomial simulation is used to simulate the epidemic process. The simulation assumes that transmission evolves in discrete generations, with new 'infections' or colonisations in each generation, binomially distributed, with the number equal to the number of susceptibles (or uncolonised individuals) and probability of colonisation, $P = 1 - e^{-\lambda_t}$. I run 100 simulations over 2 years (105 weeks) of epidemic. The R code used is shown in the appendix to this chapter.

The simulation takes two inputs:

1. State matrix S(t): an N x 2 matrix of initial states (colonised or uncolonised) at time (t). I used the observed baseline initial states.
2. Rate function: the rates at which an individual j, transitions from U -> C or C ->U, derived using the transmission model.

Simulation steps are as follows:

1. Compute rates
   $\overrightarrow{\lambda_t} = \lambda\big(t; S(t)\big)$ $\quad$ $\overrightarrow{\lambda_t}$ is a vector of colonisation rates
   $\overrightarrow{\gamma_t} = \gamma\big(t; S(t)\big)$ $\quad$ $\overrightarrow{\gamma_t}$ is a vector of decolonisation rates

2. Compute transition probabilities, given (1)
   $\vec{P}_{UC} = 1 - e^{-\lambda_t}$ $\quad$ $\vec{P}_{UC}$ is a vector of colonisation probabilities
   $\vec{P}_{CU} = 1 - e^{-\gamma_t}$ $\quad$ $\vec{P}_{UC}$ is a vector of colonisation probabilities

3. Compute new events
   $\vec{x}_{UC_t} \sim \text{Binom}\,(\vec{U}(t), \vec{P}_{UC})$ $\quad$ $\vec{x}_{UC_t}$ is a vector of new colonisation events
   $\vec{x}_{CU_t} \sim \text{Binom}\,(\vec{C}(t), \vec{P}_{CU})$ $\quad$ $\vec{x}_{CU_t}$ is a vector of new decolonisation events

4. Update vectors of colonised and uncolonized individuals at the next time step (t+1), based on (3).

5. Move forward one time step, t=t+1

6. Go to (1)

7.3.3 Model assumptions

The models assume density-dependent transmission within the household (i.e. contact rate and therefore transmission between individuals, depends on population density within the house). This is a plausible assumption, since it is likely that contact rates will be higher in more crowded households.

Conversely, the models assume frequency-dependent (density independent) global transmission (i.e. that between household contact rates are independent of population density). It is plausible that people who do not live in the same household are not in sufficiently close contact, for population density to be important.

The model makes a homogeneous mixing assumption[238] – i.e. assumes that individuals mixing within households as described by the mixing matrix, and individuals mixing between households, do so homogenously and at random.

In the U-C-U model, it is assumed that once colonised, an individual cannot be re-colonised and remains in the colonised state for a fixed mean time period, $(1/\gamma)$. Individuals were assumed to remain colonised for an average of 14 days, based on data from a previous household study in Blantyre (J.Lewis, personal communication), giving $\gamma = 1/14$. It is also assumed that household size is constant over time. In the model simulation, a time-step of one week was used because it is unlikely that a colonisation/decolonisation transition would occur at a faster rate[48].

Finally, the model assumes that everyone in each household gave a sample and does not account for any individuals who were in the house but declined to participate in the study.

7.3.4 Parameter selection and model fitting

Model calibration was performed heuristically by manual exploration of the parameter space and visual fitting of model parameters to give a simulation output close to the observed distribution of household prevalences. This process is described in more detail as follows.

First, the prevalence of 3GC-R colonisation in each household was estimated from the observed data, by dividing number of samples which were positive for 3GC-R by the total number of samples collected from each house. Prevalence was calculated at each timepoint (day 0, day 30, day 90, day 180) and displayed as histograms. The Kolmogorov-Smirnov test, a non-parametric test of equality of continuous cumulative distribution functions, was used to compare the prevalence distributions between dry and rainy season.

Plausible ranges for parameter values were selected, using an argumentation approach, and used in the model simulation as described in Section 7.3.2. The output from the model simulations was used to estimate household 3GC-R prevalence data for each household, over the two-year simulated epidemic.

Using the model simulations, a mean simulated 3GC-R prevalence was calculated for each household at two time-points, one during rainy season

and one during dry season. Mean simulated prevalence histograms were plotted for each season and visually compared to the true household prevalence distributions.

7.3.5 Sensitivity analysis

I carried out a sensitivity analysis to explore different parameter set combinations for each model. This analysis is described further in Section 7.4.3 below.

7.3.6 Computational analysis

Analysis was carried out in R.Studio. The R code used is shown in the appendix to this chapter.

7.4 Results

7.4.1 Household prevalence from observed data

Figure 7.3 shows the household prevalence distributions of 3CG-R *E. coli* colonisation, in all households at each timepoint in the study. Figure 7.4 shows the dates of sample collection for each follow up visit. Inspection of the histograms in Figure 7.3 shows more probability mass at higher prevalence values in the Day 180 plot than the plots from other timepoints and Figure 7.4 shows that these Day 180 samples were predominantly collected in rainy season.

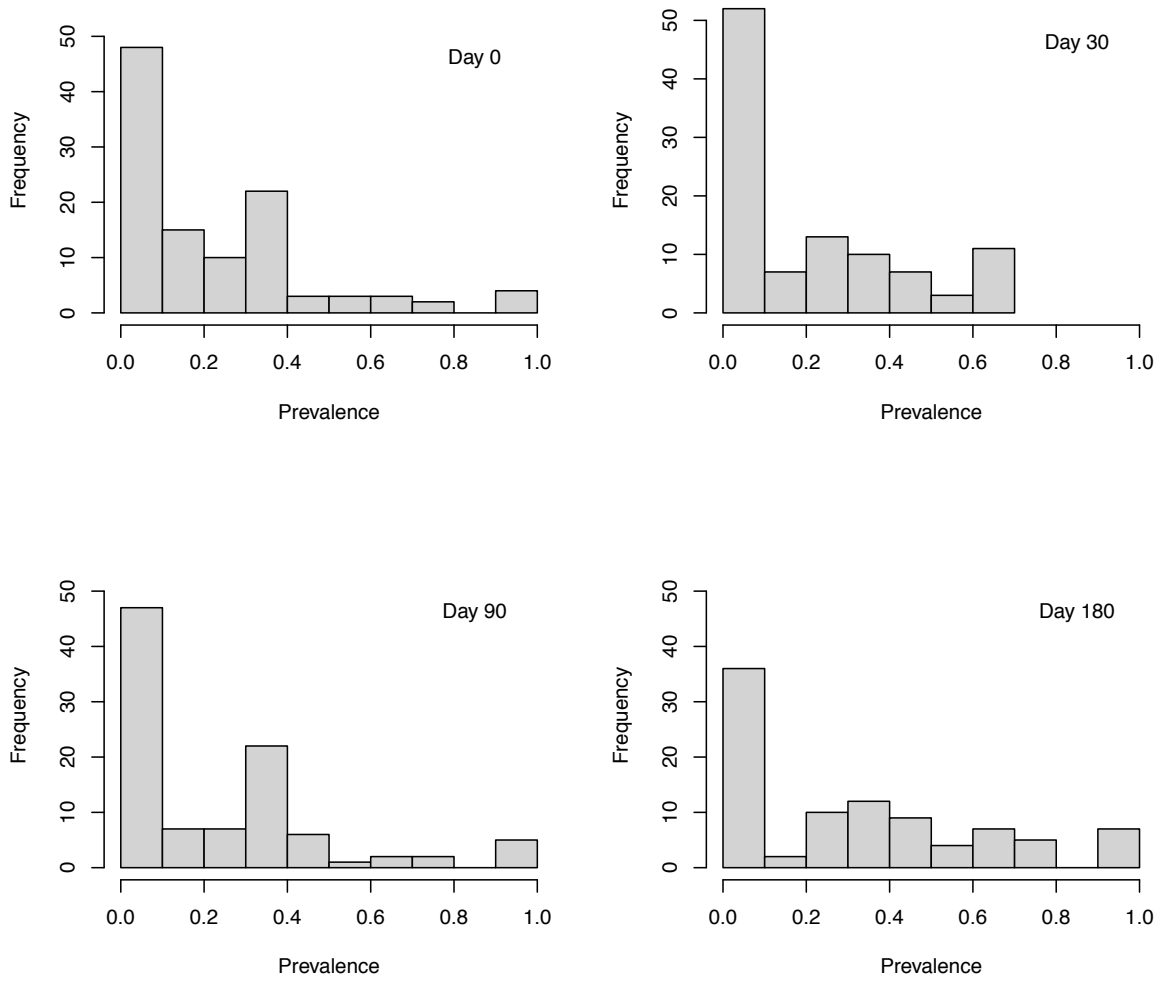


Fig 7.3: Household prevalence of 3GC-R carriage by study visit

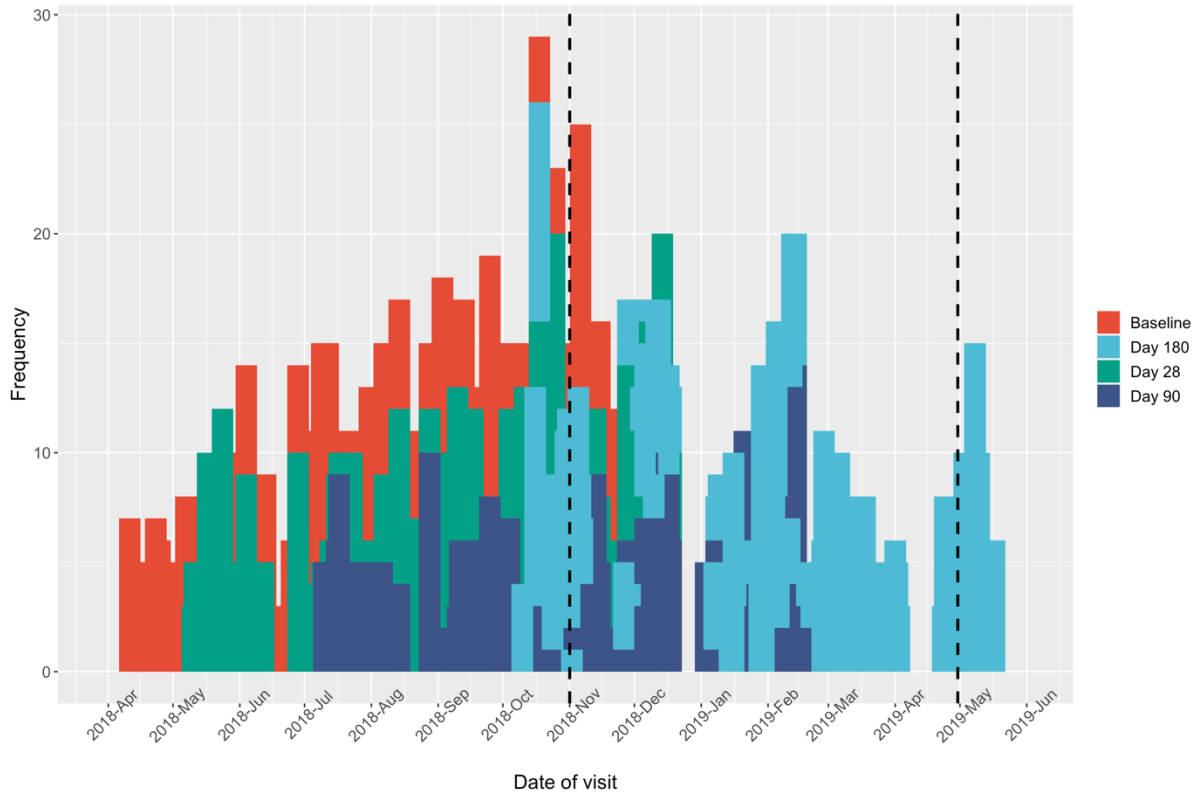


Fig 7.4: Dates of stool sample collection stratified by study visit. Rainy season is indicated by the two dotted lines and was defined as the time period between 1st November and 30th April.

The results of the Kolmogorov-Smirnov tests demonstrate strong evidence that the distribution of household prevalence at the Day 180 visit is right skewed relative to that at baseline (Table 7.1, $p<0.01$). The distribution of household prevalence at the Day 30 and Day 90 visits, however, are not significantly different from baseline (Table 7.1). It is likely that this difference has occurred because of the seasonal differences in sample collection at each study visit time point, as shown in Figure 7.4 and as described in Chapter 6.

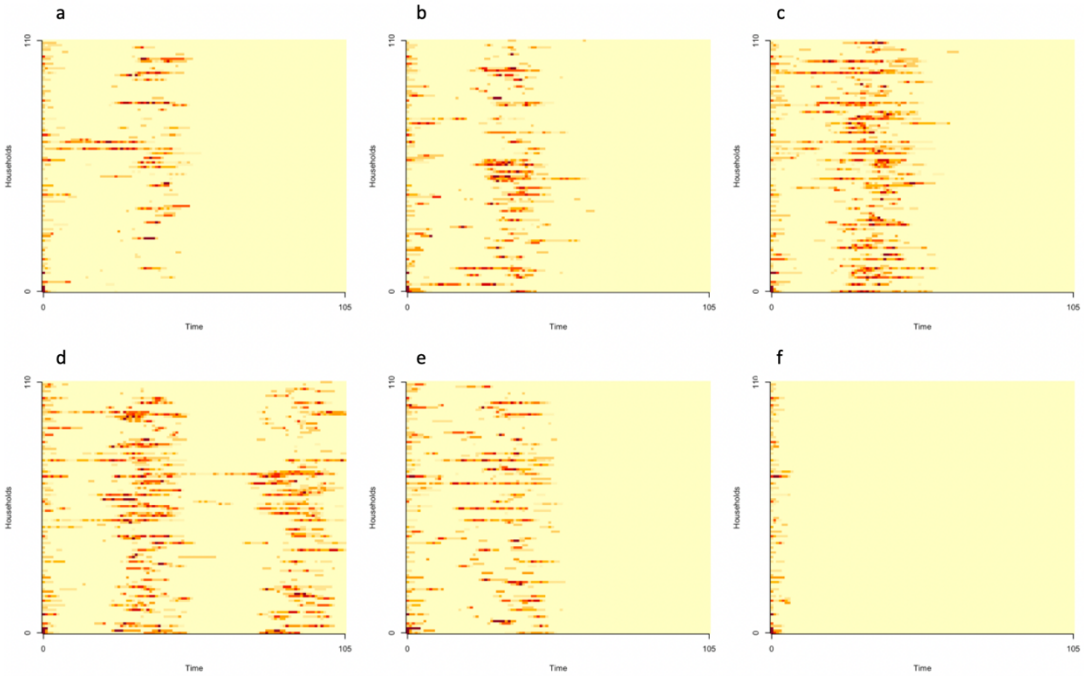
Table 7.1: One-sided Kolmogorov-Smirnov tests comparing household prevalence distributions of 3GC-R *E. coli*

Timepoints	Test statistic	p-value
Day0/Day30	0.012776	1
Day0/Day90	0.034859	0.9643
Day0/Day180	0.11652	0.004686

7.4.2 Household prevalence from simulations

Figure 7.5 shows a heatmap visualisation of the simulated household prevalence data from 6 randomly selected simulations, generated using Model 1 (Figure 7.5A) and using Model 2 (Figure 7.5B). Using Model 1, in which seasonality is applied to both within and between household transmission, clear seasonal prevalence peaks are seen (Figure 7.5A). Stochastic fade-out, which occurs when, by chance, everyone in the population becomes uncolonized such that both transition rates, U->C and C->U fall to zero (i.e. an absorbing state), is more common in Model 1 than Model 2. Conversely, using Model 2, which applies seasonality to between-household transmission only, colonisation is much more sustained throughout the year. We see that households which have a high prevalence at the end of the wet season are more likely to remain high prevalence across the dry season. These households are then responsible for "re-igniting" between-household transmission at the start of the next wet season.

A



B

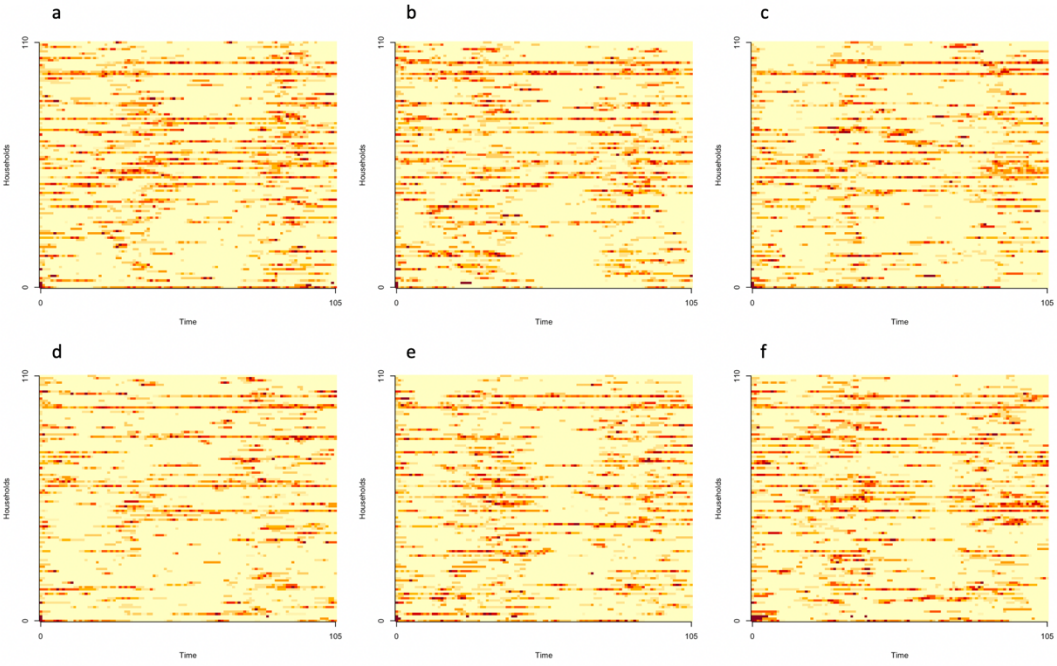


Figure 7.5: Heatmaps illustrating household prevalence of 3GC-R colonisation over 105 weeks of simulated time. Figures A-F are 6 randomly selected simulations (from total of 100). Each row represents a household. Coloured pixels correspond to the prevalence values using a scale (dark red = highest prevalence, pale yellow = prevalence of zero). In (A), plots a-e show clear seasonal peaks. Many households fail to become recolonised by the second rainy season. Figure F shows stochastic fade-out – in this simulation, the outbreak ends quickly in all households. In (B), household prevalence is more homogenous throughout the two years. Most households become recolonised by the second rainy season. Some households keep colonised throughout the 2 years whereas some lose and gain colonisation

Figure 7.6 and Figure 7.7 show model simulation outputs generated using four different parameter sets, using Model 1 and Model 2, respectively.

In all cases, the observed prevalence data are much more dispersed when compared to the simulated prevalence. This is expected, since the models do not currently account for the individual level heterogeneity which is present in the observed data. In Figure 7.6, we see that the deep troughs in dry season, result in loss of colonisation in almost all households (t=60, blue histograms). In plot (a), where β_G and β_{HH} are the same size, transmission re-ignites in rainy season, though too few households are becoming uncolonized when compared to the observed data. In plots (b), where β_{HH} is considerably smaller than β_G, stochastic fade-out occurs and there is no re-ignition of colonisation in the subsequent rainy season. In plot (c), where β_{HH} is slightly higher than β_G, transmission is re-ignighted in some individuals but household prevalence remains very low. In plot (d), β_{HH} is approximately 1.5 times larger than β_G, and some households do become re-colonised, leading to a better visual fit with the observed data, but with more probability mass at lower prevalence than the observed data.

In Figure 7.7, we see reduced amplitude of seasonal peaks when compared to Figure 7.6 overall. As with Model 1, in plots (b) and (c), we see loss of colonisation of almost all households in dry season, with failure to re-ignite in wet season. In plot (a), where β_G and β_{HH} are the same size, as with Model 1, transmission re-ignites in rainy season, but again too few households are becoming uncolonized when compared to the observed data.

In plot (d) β_{HH} is approximately 1.5 times larger than β_G, and the best overall visual fit with the observed data is seen.

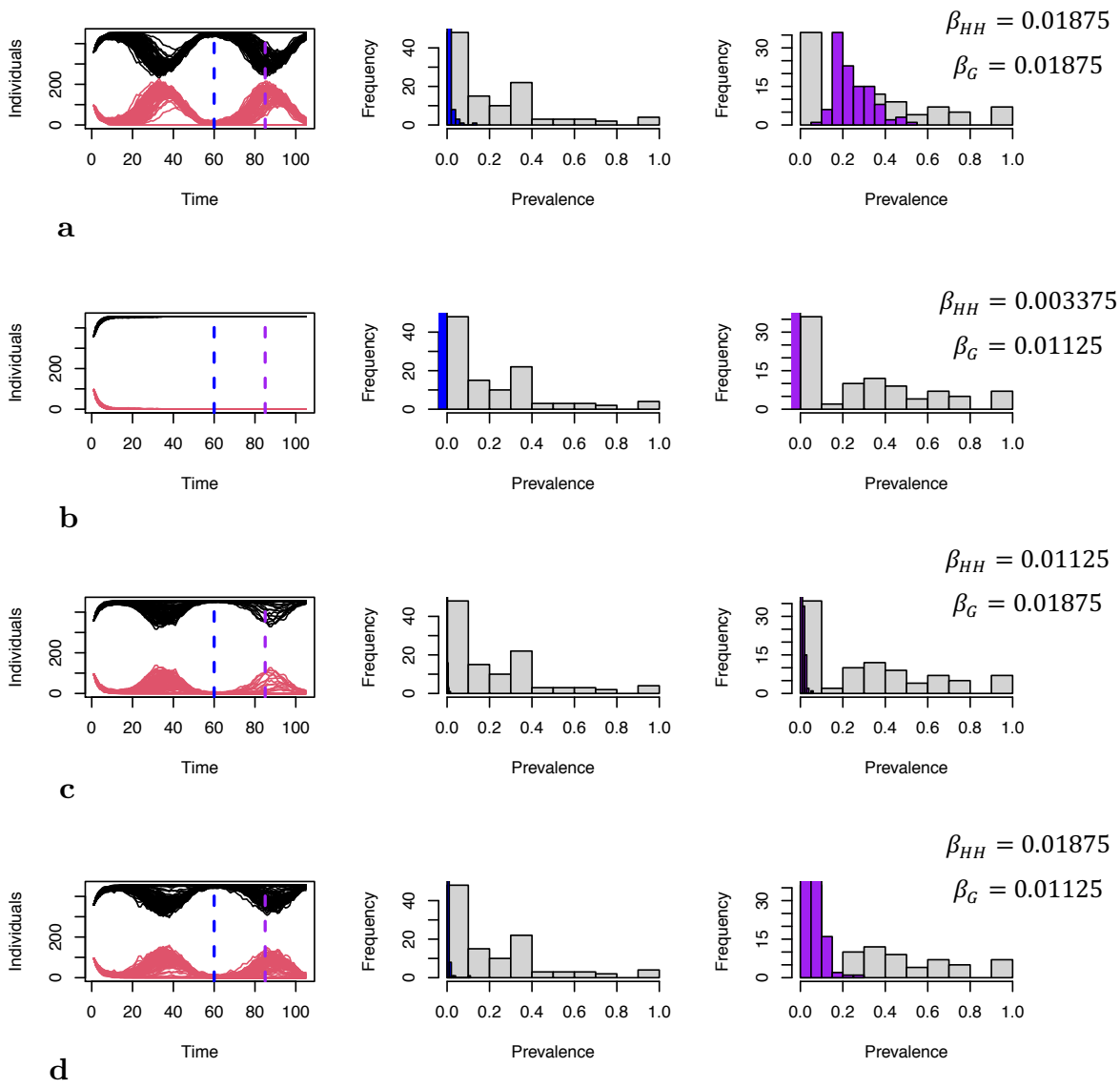


Figure 7.6: Model simulation outputs generated using four different parameter sets, using Model 1. The line-plots show absolute numbers of colonised (red) and uncolonised (black) individuals, over time. The adjacent histograms summarise household prevalence at two time points: dry season (blue) and rainy season (purple), shown alongside prevalence histograms of the 'true' observed data from each season.

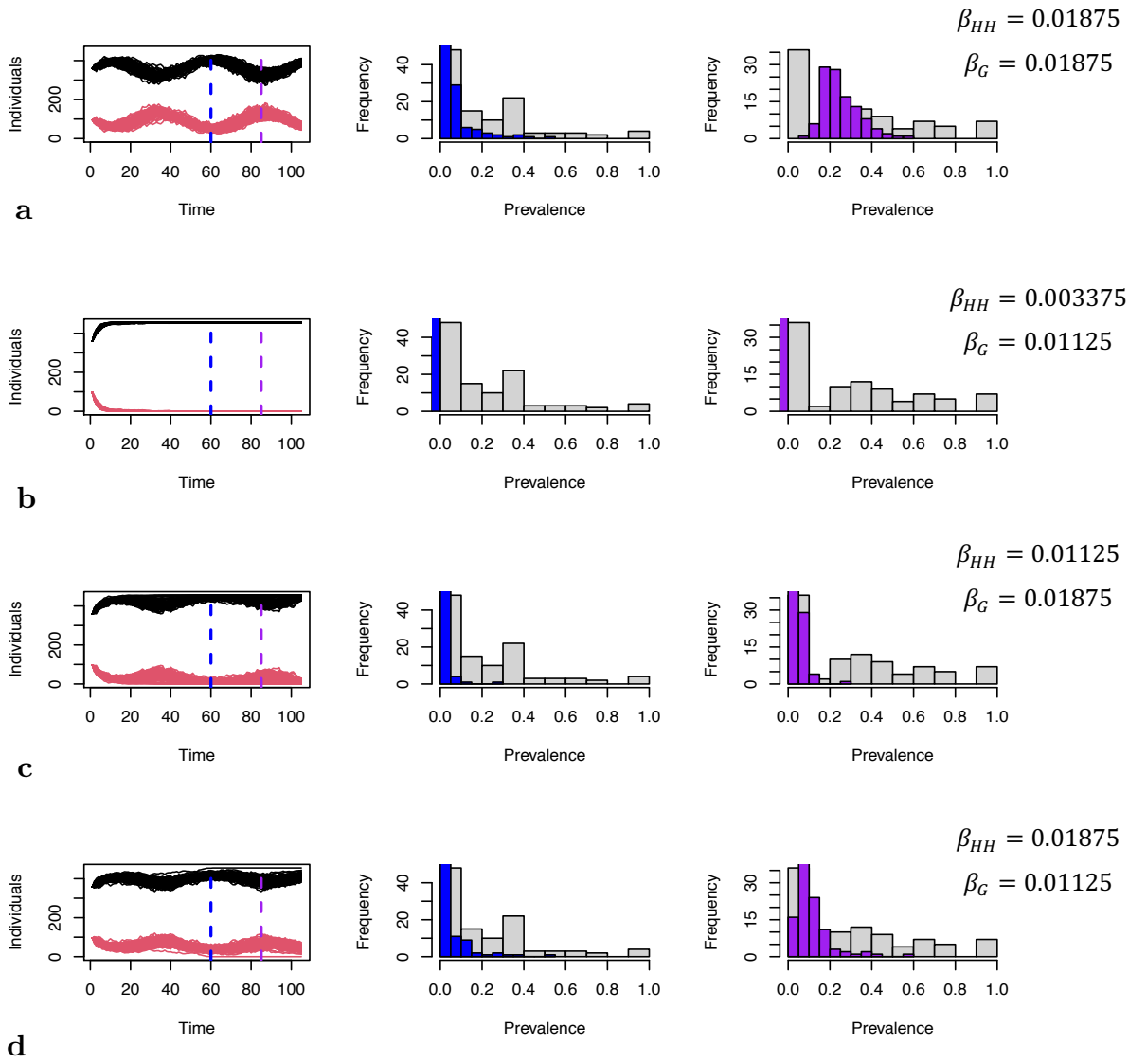


Figure 7.7: Simulation outputs generated using four different parameter sets, using Model 2

7.4.3 Sensitivity analysis

In the above section, I show that for one particular parameter combination, there is a high probability that stochastic fade out occurs by the end of the simulated time period. In this section, I explore this finding using more parameter set combinations to determine household 3GC-R persistence, where:

Probability (Pr) of persistence, is 1-Pr(C(T)=0)

First, β_{HH} and β_G were varied between 0.01 and 0.3 and the seasonal amplitude varied between 0 and 1. A parameter grid of all combinations of these three parameters was then generated. Two-hundred model simulations were run for each parameter set in the grid, using Model 1 and Model 2.

Figure 7.8 shows heatmaps of colonisation persistence as a function of the parameters using Model 1 and figure 7.9 shows these heatmaps generated using model 2. In the Model 1 plots, we see persistence is more sensitive to changes in the seasonal amplitude than in Model 2 and that overall, there is more persistence in Model 2 than in Model 1. In both sets of plots, we see that persistence is highly sensitive to changes in the β_{HH}, but that changes in β_G have much less effect. This is explored further in Figure 7.10, which shows a line plot of persistence for changing values of β_G, using $\beta_{HH} = 0.015$. In these plots, we see that, in Model 2 (Figure 7.10 (B), for a fixed middle value of β_{HH}, a high β_G is needed to avoid stochastic fade out and in Model 1, this value of β_G is lower (Figure 7.10 (A)

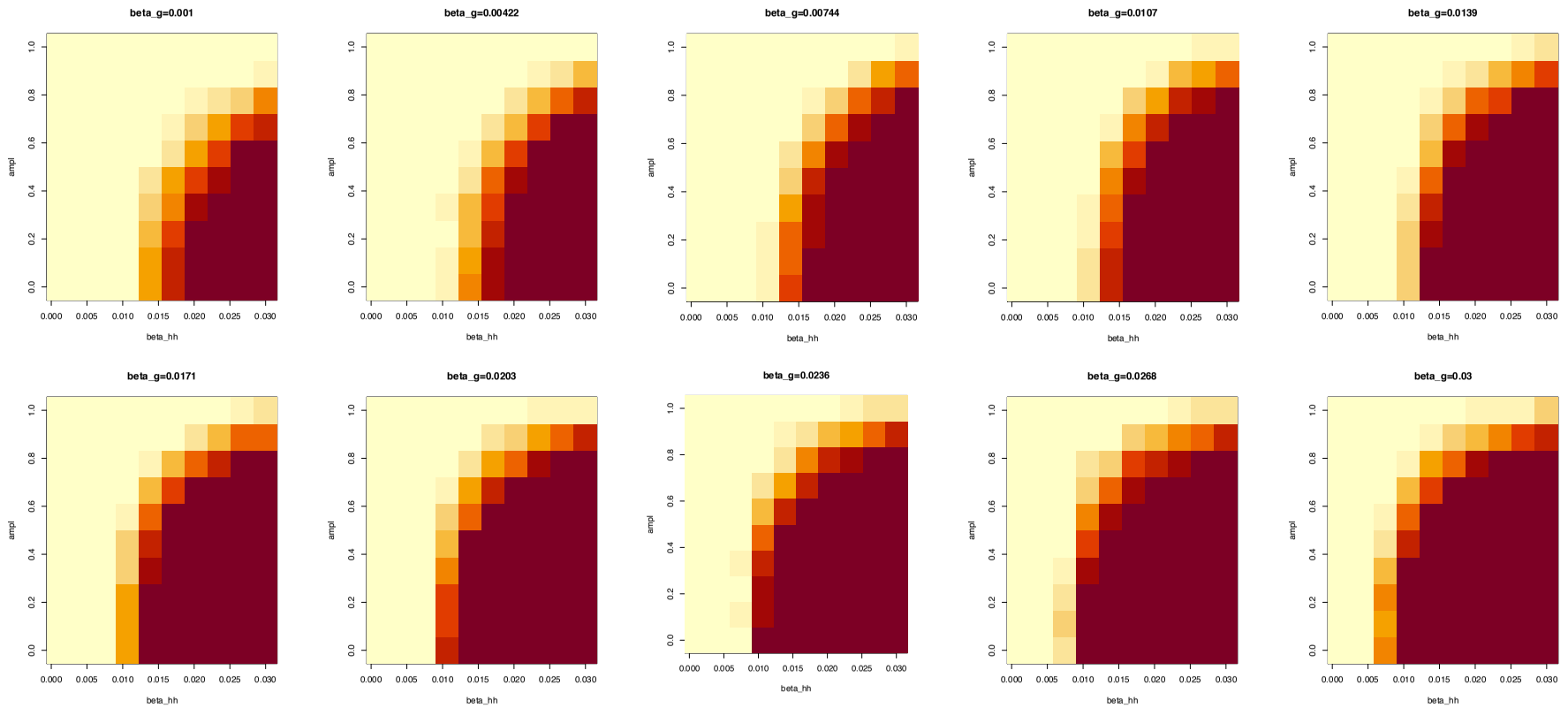


Figure 7.8: Heatmaps showing colonisation persistence for ten different parameter sets, using Model 1. Coloured pixels correspond to the proportion of simulations in which at least one individual was colonised at the end of T, using a scale using a scale (dark red = highest proportion, pale yellow = proportion of zero). beta_hh = household transmission parameter, beta_g = global transmission parameter

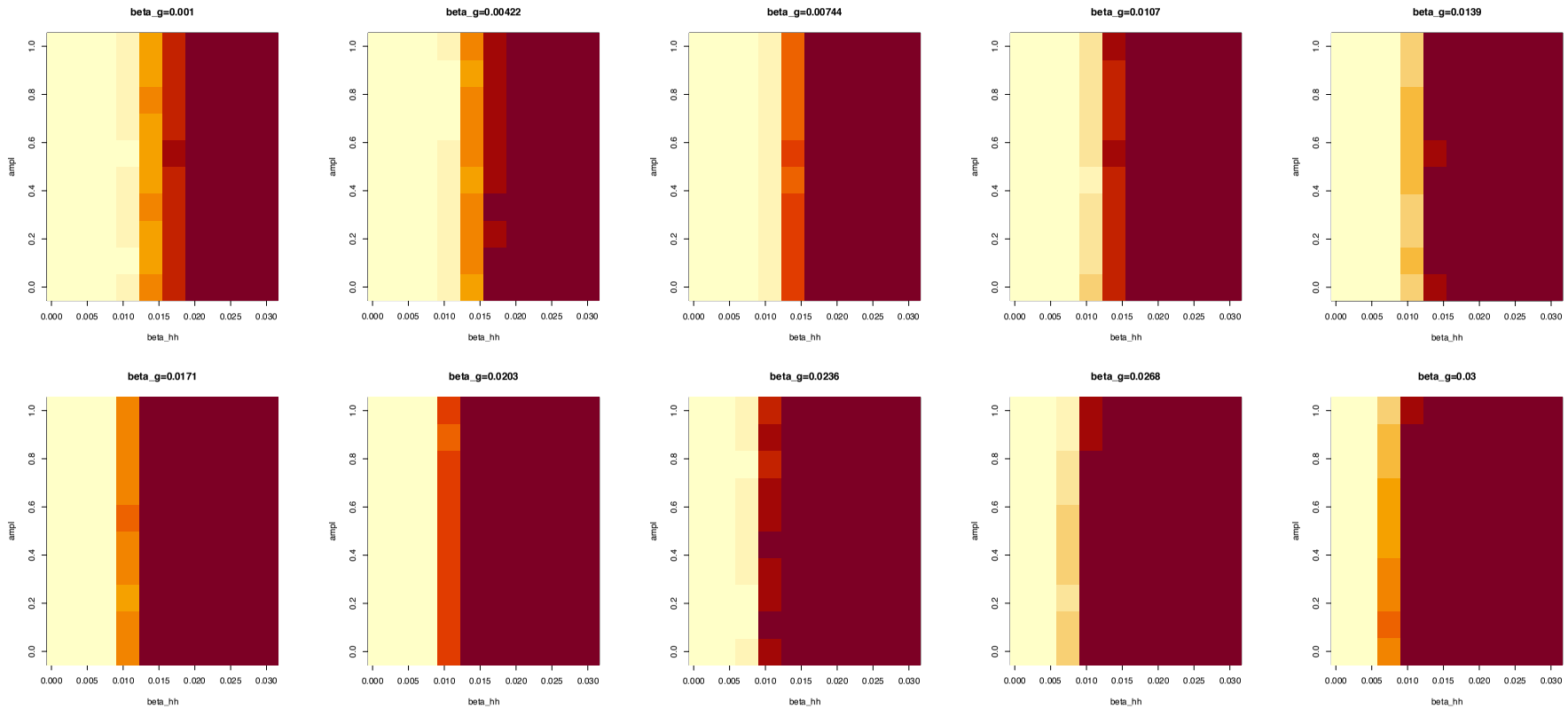


Figure 7.9: Heatmaps showing colonisation persistence for ten different parameter sets, using Model 2. Coloured pixels correspond to the proportion ofsimulations in which at least one individual was colonised at the end of the two year simulated time period, using a scale using a scale (dark red = highest proportion, pale yellow = proportion of zero). beta_hh = household transmission parameter, beta_g = global transmission parameter

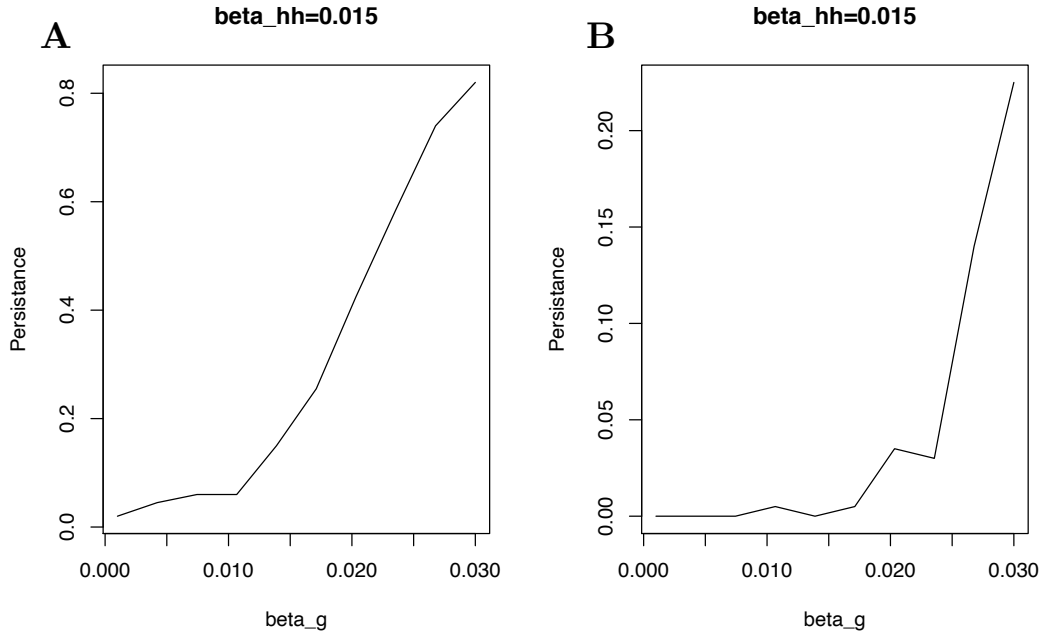


Figure 7.10: (A) shows persistance probability for varying values of β_G, using Model 1 and (B) Model 2. beta_g = global transmission parameter, beta_hh = household transmission parameter

7.5 Discussion

In this chapter, I have developed a framework for a dynamical transmission model with network structure, that can be used to infer the transmission dynamics of 3CG-R *E. coli* amongst adults and children in a community setting in urban Blantyre. The models incorporate parameters for within and between household transmission and can be therefore be used to explore the relative importance of infectious person-to-person spread, versus environmental reservoirs. I have fitted the models to household prevalence data derived from the community cohort described in Chapter 6, using manual parameter exploration and visual fitting. Results from this preliminary analysis suggest that the within household network may be a key driver of transmission, against which interventions should be targeted.

The models I have used describe two strongly connected networks: within and between household. On the basis of the homogeneous mixing assumption, which states that each individual outside the household has a small and equal chance of coming into contact with any other individual, these two networks are also weakly connected to each other, so that the relative sizes of the parameter pair are important[239]. In the simulations for both Models 1 and 2, I show that if the within household transmission parameter is very small in relation to the between household parameter, transmission cannot be sustained. Conversely, if household transmission is high, even a small amount of environmental transmission can sustain the epidemic. This is corroborated by the sensitivity analysis, which demonstrates that persistent colonisation is highly sensitive to changes in household transmission, but much less so for global transmission. If we think of this in terms of targeting interventions, the models suggest that if we target and 'clean' the environmental reservoir of 3CG-R, we would sever connections between the households but remain with infected cliques of transmission circulating within a small number of houses. If, however, we 'clean' the households, the epidemic might die out, because environmental transmission alone cannot sustain it.

The next step then, is to understand this within household transmission in more depth. To do this, the models will need to be refined, with incorporation of individual level parameters, which may be time-varying, such as antibiotic use (including CPT) and healthcare exposures, or time-fixed, such as age and sex. Indeed, one of the limitations of the models in their current form, is the assumption that all individuals are the same,

which has resulted in fairly homogenous household prevalence data from the simulations.

In Chapter 6, I showed that sample collection during rainy season was a significant risk factor for 3GC-R colonisation. In this chapter, I examine this effect in more detail, using models which look at the impact of season on within and between household transmission. In the model which applies seasonal peaks and troughs to the household and the external environment, we see transmission effectively fading out in the dry season, to the extent that it cannot be re-ignited during the subsequent rains. When we then remove these dry season troughs from the household (Model 2), transmission becomes far more sustained, in particular within some households, leading to persistent transmission throughout the year. This latter model likely represents a more plausible scenario. The household environment may be relatively protected from the effects of the change in season, and within household pockets of infection persist throughout the period when the environment is not conducive to 3GC-R spread (e.g. dry weather). In the wet season, 3CG-R is homogenous, with many households 'infected', but in the dry season, this homogenous transmission becomes segregated into the 'protected' household network, which then sustains the epidemic throughout the year. This is further evidence that the within household network needs to be explored as an area to target interventions.

7.5.1 Limitations

The models and analysis in this chapter have a number of limitations that will need to be addressed as the models are refined. Mainly, I have used a relatively rudimentary method of model fitting, estimating parameters

manually through comparison of prevalence outputs. This is useful as an initial approach to testing these models, given the complex computational requirements of gold standard methods of model estimation, such as likelihood-based inference, but future work should involve a more comprehensive exploration of the parameter space, together with measures of uncertainty. The question of what values of β_{HH} and β_G would produce an epidemic most consistent with our observed data, can be tackled with a likelihood function, $f(x|\theta)$, i.e. the likelihood of observing our data (x), given the set of parameters (θ). Once the likelihood function is derived, we can then take a Bayesian or a frequentist approach to deriving $P,(\theta|x)$, the probability of the parameter values, given the data[236, 240, 241].

The assumptions I have made in the analysis need to be explored further in future iterations of the models. The assumption of random mixing of individuals within and between households, is unlikely to be realistic. Far more likely, is that within urban Blantyre, different mixing rates are experienced between population subgroups (e.g. children versus adults) and future models should ideally incorporate additional transmission matrices representing, for example, school and workplace networks. To do this well, we should first gather more detailed information on social networks within Blantyre, potentially via diary based studies which ask individuals to recall their contacts – a technique frequently used for studies of respiratory infection[242-245] and HIV transmission[246].

The assumption of a 14-day decolonisation rate is based on very limited data from a single study of hospitalised patients in Blantyre. More data is needed to inform this parameter assumption and the effects of varying the

decolonisation rate should be explored. In order to further characterise decolonisation rates, studies would need to undertake intensive sampling at much higher frequencies than those used in my cohort. As I discuss in Chapter 1, there is currently only one published study which attempts to overcome the issues of interval-censored data experienced by studies which have sparse sampling intervals, using daily sampling of participants (travellers to Laos) over a three-week period[61]. The status of individuals varied daily, suggesting multiple transient colonization events, lasting a few days at most. Intensive sampling work from community cohorts in Blantyre could better inform the models in this chapter.

The models I have used assume that 3GC-R arises exogenously when in fact it can occur endogenously from within the host. Future iterations of the models could incorporate an additional term which models this endogenous change and this may allow de novo resistance to restart transmission from otherwise extinct states.

I have assumed that the diagnostic test used to culture 3GC-R *E. coli* (selective Chromogenic agar) is 100% sensitive and specific, such that the 3GC-R status of an individual represents the true state. Future models should account for uncertainty in the diagnostic testing process, as well as the different methods used for culture (rectal swab versus whole stool). Furthermore, I have modelled phenotypic 3CG-R status from individuals, and future models should aim to incorporate data from WGS isolate typing to provide greater resolution to discriminate between different *E. coli* strains and provide further insight into potential transmission routes.

7.6 Conclusions and future work.

In this chapter I have developed dynamical transmission models, which are a feasible representation of the process of community level transmission of 3CG-R amongst adults and children in urban Blantyre.
Although I use rudimentary fitting methods, the models provide insight into the into the relative importance of within verses between household transmission, describing a potentially important role for within household reservoirs.

Ultimately, the aim of future work should be to develop a comprehensive model which can be used to determine detailed risk factors for colonisation and identify where transmission pathways can be interrupted. The next iterations of the models in this chapter, should therefore incorporate individual-level parameters and social network information and should be fit using robust parameter inference. Future studies should aim to gather detailed contact information that can be used to parameterise social networks in these models and should make use of WGS to capture organism diversity and provide further insight into the dynamics of 3GC-R *E. coli* carriage.

7.7 Appendix

```
## Packages
abind_1.4-5
raster 3.0-7
tidyr 1.1-3

#       Discrete-time simulation
#
# Simulates a discrete time epidemic
#
# @param initial_state a matrix of number of individuals by number of
states
# @param rate_fn a callable taking the state matrix and returning a
matrix of number of individuals by number of transitions.
# @param num_steps number of timesteps required
# @param start start time
# @param time_delta the size of the timestep (default 1)
# @param stoichiometry matrix giving state update rules for each
transition.
# Rows represent transitions, columns represent states.  Default UCU
model.
# @returns matrix of state at each timepoint

simulate = function(initial_state,
                    rate_fn,
                    num_steps,
                    start=0,
                    time_delta=1,
                    stoichiometry=matrix(c(-1, 1, 1, -1), ncol=2)) {
  state = initial_state
  result = array(dim=c(dim(initial_state), num_steps+1),
                 dimnames=c("individual"=NULL, "state"=NULL,
"step"=NULL)
  )
  result[,,1] = initial_state
  times = seq(from=start, length.out=num_steps, by=time_delta)
  for(step in 1:num_steps) {
```

```
    transition_rates = rate_fn(times[step], state) # step 1
    transition_probs = 1 - exp(-transition_rates*time_delta)# step 2

    # Compute new events
    new_events = matrix(rbinom(length(transition_probs),# step 3
                               size=state,
                               prob=transition_probs
    ),
    ncol=ncol(state)
    )
    state = state + new_events %*% stoichiometry # step 4, apply  states
to events to update new states at each time point
    result[,,step+1] = state # stores state in 3D array [N, S, T700+900]
  }

  result
}
```

UCU model

```
#########################################
#
# Covariate information
#
# HH matrix (within house prevalence)
# Season
#########################################

source("simulator.r")

# Set up  household matrix
N = 455
Kmembership = t(sapply(unique(df0$hid), function(hh)
as.integer(df0$hid==hh))) #hh X ind
network = t(Kmembership)%*%Kmembership
diag(network) = 0  # People cannot infect themselves
hh_matrix<-network

# Set up initial conditions
```

```
# 1 if in state, 0 otherwise
drop_na(df0,result)->df0
C0<-df0$result
U0 <-abs(C0 - 1)

#' Transition rates
#'make_transition_rates  - function makes transition rates to pass to
simulation
#The simulation takes a function, which given current state of the
population, returns all transition rates

#' Function to build transition rates
make_transition_rates = function(params, covariates) {

  rate_fn = function(t, states) {
    # U->C
    season_effect = 1+params$ampl * -cos(t*2*pi/365)
    household_rate = params$beta_hh * (covariates$hh_matrix %*%
states[,2])
    global_rate = params$beta_g * sum(states[,2]) / nrow(states)
    colonisation_rate = (season_effect) * (household_rate + global_rate)

    #colonisation_rate = household_rate + (season_effect *global_rate)



    # C->U
    decolonisation_rate = params$gamma
    cbind(colonisation_rate, decolonisation_rate)
  }
  return(rate_fn)
}
par(mfrow=c(4,2))
# e.g. params = list(beta_hh=0.025*.75, beta_g=.025*.75, gamma=1/14,
ampl=0.7)



# Run simulation
## Constants
initial_state = cbind(U0, C0) # Matrix of uncol and col individuals
#params = list(beta_hh=0.75, beta_g=.5, gamma=1/14, ampl=0.7)
```

```
covariates = list(hh_matrix=hh_matrix)

## Make rate Function
≈
rate_fn = make_transition_rates(params, covariates)
sim = simulate(initial_state=initial_state,
               rate_fn=rate_fn,
               num_steps=365*2/7,
               time_delta=7.)

state_size = apply(sim, c(2, 3), sum)


nsims = 100
sims=replicate(nsims,simulate(initial_state=initial_state,
                              rate_fn=rate_fn,
                              num_steps=365*2/7,
                              time_delta=7.))


state_size = apply(sims, c(2,3,4), sum)

plot(state_size[1,,1], col=1, type='l', ylab="Individuals", xlab="Time",
ylim=c(0, 455))
lines(state_size[2,,1], col=2)
for(i in 2:nsims) {
  lines(state_size[1,,i], col=1)
  lines(state_size[2,,i], col=2)
}


# get  prevalence for households at t=85 and plot it
colonised = sims[,2,,]

prevalence = abind(lapply(1:dim(colonised)[3],
                          function(s)
                            abind(
                              lapply(85,
                                     function(t)
                                       Kmembership %*% colonised[, t, s] /
rowSums(Kmembership)),
```

```
                              along = 2
                            )), along = 3)
hist(Kmembership %*% df0$result / rowSums(Kmembership),
xlab="Prevalence",main="")#real data
hist(apply(prevalence, c(1,2), mean), add=TRUE, col = 'blue')# model
data




# function to optimise parameters,fit to prevalence at t=85
initial_state = cbind(U0, C0)
covariates = list(hh_matrix=hh_matrix)
opt_fn = function(par) {
  params = list(beta_hh=par[1],
                beta_g=par[2],
                gamma=1/14,
                ampl=0.6)
  nsims = 100
  sims=replicate(nsims,simulate(initial_state=initial_state,
                                rate_fn=rate_fn,
                                num_steps=365*2/7,
                                time_delta=7.))
  colonised = sims[,2,,]
  prevalence = abind(lapply(1:dim(colonised)[3],
                            function(s)
                              abind(
                                lapply(85,
                                       function(t)
                                         Kmembership %*% colonised[, t, s] /
rowSums(Kmembership)),
                                along = 2
                              )), along = 3)
  mean_prevs = apply(prevalence, c(1,2), mean)
  ks.test(hh_prev, mean_prevs)$statistic
}
optim(par=c(0.025*0.75, 0.015*0.75), fn=opt_fn)
```

```
# images of simulated prevalence

prevalence = abind(lapply(1:dim(colonised)[3],
                          function(s)
                            abind(
                              lapply(1:dim(colonised)[2],
                                     function(t)
                                       Kmembership %*% colonised[, t, s] /
rowSums(Kmembership)),
                              along = 2
                            )), along = 3)

x3 <- sample(1:100, 6, replace = T)# random 6
par(mfrow=c(2,3))
image(t(prevalence[,,60]), axes=FALSE, xlab="Time",ylab="Households")
at.seq = seq(from=0,to=1,length.out=2)
label.seq = c(0,105)
axis( side=1, at=at.seq, labels=label.seq )

label.seq = c(0,110)
axis( side=2, at=at.seq, labels=label.seq )


image(t(prevalence[,,60]), axes=FALSE, xlab="Time",ylab="Households")
at.seq = seq(from=0,to=1,length.out=2)
label.seq = c(0,105)
axis( side=1, at=at.seq, labels=label.seq )

label.seq = c(0,110)
axis( side=2, at=at.seq, labels=label.seq )

at.seq = seq(from=0,to=1,length.out=2)
label.seq = c(0,105)
axis( side=1, at=at.seq, labels=label.seq )

label.seq = c(0,110)
axis( side=2, at=at.seq, labels=label.seq )
box()

# sensitivity analysis
```

```
beta_hh <- seq(0.001,0.03,len=10)
beta_g <- seq(0.001,0.03,len=10)
ampl <- seq(0,1, len=10)


paramgrid <-expand.grid(beta_hh=beta_hh,beta_g=beta_g,ampl=ampl)

#loop to run sims for each parameter row
nsims=200# increase
is_not_extinct=numeric(nrow(paramgrid))#container for results
for(i in 1:nrow(paramgrid) ){
  params=as.list(paramgrid[i,])
  params$gamma=1/14
  print(params)
  rate_fn=make_transition_rates(params=params,covariates =
covariates)
  sims=replicate(nsims,simulate(initial_state=initial_state,
                                rate_fn=rate_fn,
                                num_steps=365*2/7,
                                time_delta=7.))
  sims=sims[,2,105,]
  prob=mean(apply(sims,2,sum)>0) # returns scalar of proportion of sims
that have at least one infected at end of time
  is_not_extinct[i]=prob
}
```

Chapter 8

Conclusions and future work

8.1 Summary of findings

I have presented the findings from two longitudinal cohort studies based in Malawi. The hospital cohort had the broad aim of understanding the outcomes of 3GC-R BSI on patients and the household cohort aimed to describe risk-factors for 3GC-R colonisation. The hospital cohort was a clinical study of patients with bloodstream infection caused by Enterobacterales or *Acinetobacter* spp., the results of which are presented in Chapters 3, 4 and 5.

In Chapter 3, I described the clinical and microbiological features of the 336 recruited patients and used multivariable models to investigate risk factors for 3GC-R BSI. *E. coli* and *Klebsiella* spp. were the most frequently isolated pathogens and 72% of all included organisms were 3GC-R. Participants were young overall, with just under one-third of the cohort aged under three-months. Prior to my study, it was assumed that these BSI were predominantly community-acquired, but I show that 63% were healthcare associated and that a higher proportion of hospital onset BSI were 3GC-R than 3GC-S (36.0% vs 21.0%, $p<0.001$). A high early mortality, meant that 31.8% of patients had died before the final blood culture result was available and data from these patients were captured posthumously. Missing data were therefore a problem and so imputed datasets were generated for construction of multivariable models of risk-

factors. Although these models were imperfect, prior healthcare exposures, in particular prior surgery were associated with 3CG-R. Most infants recruited were born outside of QECH, and appeared to acquire their infections in other healthcare settings.

In Chapter 4, I showed that these BSI carried an extremely high mortality irrespective of antimicrobial susceptibility, with a CFR of 43% and that 3GC-R was associated with even higher in-hospital mortality (OR 1.85, 95%CI 1.06-3.26), greater risk of death (HR 1.51, 95%CI 1.04-2.18) and longer hospital stays (HR for discharge HR 0.70, 95%CI 0.49-0.99). Overall survival was poorest in the youngest patients but the study was underpowered to detect 3GC-R mortality associations by age-group. Notably, 20.8% of patients survived to 6-months having never received an effective antibiotic for their infection.

In Chapter 5, I showed that 3GC-R was associated with adverse health economic outcomes for the healthcare provider and the patient. I found that admissions with 3GC-R placed a substantial burden on QECH, costing an average of US$106.42 more than 3GC-S infections. Patient out-of-pocket spending was not substantially different but the average indirect costs incurred from loss of income was significantly higher in 3GC-R than 3GC-S infection, with a mean difference of US$167.36. In models adjusted for age, sex and causative organism, 3GC-R was associated with higher direct and societal costs as well as lower EQ-5D scores. I extrapolated these findings to estimate the economic burden of 3GC-R in Malawi by applying the costing outputs to projected BSI incidence and population data. I found that *E. coli* and *Klebsiella* spp. could account for an annual

health provider spend of over US$1,500,000 and that approximately 200 QALYs are lost per 100,000 Malawians, from each of 3GC-R *E. coli* and 3GC-R *Klebsiella* alone.

Analysis of the household cohort was presented in Chapters 6 and 7. This study collected stool samples from adults and children and used ESBL selective Chromogenic agar to culture for 3GC-R *E. coli*. In Chapter 7, I showed that the prevalence of 3GC-R carriage was high (28.3%, 95%CI, 23.9-33.0 at baseline) and that individuals switched status multiple times over the course of the study. I used mixed-effects models, to show that a higher carriage prevalence within the household and sample collection during rainy season were strongly associated with 3GC-R colonisation. I found no effect of short-course antibiotic use, but that a household member on CPT was associated with 3GC-R on univariable analysis.

Given the limitations of hierarchical models in describing an infectious process, I used Chapter 7 to develop a framework for two dynamical transmission models, with the aim of distinguishing the relative importance of within and between household transmission. Using a two-state model with network structure and visual fitting to observed household 3GC-R prevalence, I showed that the within household transmission network is likely to be a key driver of colonisation and suggested a number of refinements for future iterations of these models.

8.2 Conclusions and future research questions

The findings described in this thesis generate two further research questions:

1. How can the burden of AMR estimates for Malawi be improved?

The hospital cohort study in this thesis, describes detailed health and economic outcomes for one of the most commonly used and frequently last-line antibiotics in hospitalised patients in Malawi and will thus provide some of the first AMR healthcare burden estimates in sSA.

My study, however, was limited to one hospital using blood culture positivity as an entry point. A straightforward next step will be to apply the outcomes to population incidence data for Malawi, in order to generate burden estimates for the country. A deeper understanding of the burden of AMR will be developed by including other pathogens and clinical syndromes, plus larger cohorts to permit age stratification, all of which will further improve our understanding of AMR in Malawi. Ultimately clinically orientated surveillance that links routinely captured laboratory data to patient outcomes, is needed. QECH is a site for ACORN (A clinically-oriented antimicrobial resistance surveillance network), a diagnostic stewardship project which aims to move health systems to electronic record keeping for patient care as well as for laboratories, with the aim of generating prospective surveillance data on DRI burden[247].

The methodological approaches used by AMR burden projects need to be harmonised so that useful comparisons within and between countries can be made. The WHO GLASS has recently published a framework for a standardised methodology which aims to ensure countries can generate comparable attributable mortality data from surveillance and research studies[142]. The methods I have used are in-line with the GLASS recommendations, which include a prospective cohort design and recruitment of a comparator group which can be drug-susceptible BSIs, non-infected controls or both.

Even with robust tools and standardised methods, determining attributable mortality from DRI remains a challenge. The prospective nature of my study, allowed for detailed clinical assessments of each recruited patient and thus enabled me to observe that some patients with 3GC-R BSI survive without an effective antibiotic. In some cases, it was my clinical judgement that the blood culture result was unlikely to be contributing to the patient's clinical condition, a judgement that was also made by other physicians attending the patient. The capacity to make these judgements is particularly important for successful antimicrobial stewardship, as carbapenems become increasingly available in Malawi. It would be unusual to classify Enterobacterales or *Acinetobacter* as a contaminant and as such, they are routinely included in surveillance studies without further consideration. This is a reasonable approach and is the one I have taken in Chapter 4, where all results are included in the analysis. It is possible, however, that the approach should be more nuanced and I propose a methodology for classifying the impact of each positive BC on a patient, with incorporation of these into future analysis.

The methods for this process are described fully in my published study protocol [145]. Briefly, in collaboration with a panel of locally experienced microbiologists, I have developed a set of categories which can be applied to the patients I have recruited, to enable classification into definite, probable or possible Gram-negative sepsis, occult or transient bacteraemia and definite or probable contaminant[145]. I have piloted these classifications, by presenting a clinical vignette for each adult patient to the expert panel, and asking them to anonymously classify each patient. Any discrepancies were resolved by consensus discussion. Table 8.1 shows some example vignettes as presented to the panel, together with the classification decided.

Table 8.1: Participant vignettes: three example participants discussed at the consensus meeting, shown with final decisions on patient classification. Taken from[145]	
Participant 1 *E. coli* S: ceftriaxone, chloramphenicol, gentamicin, ciprofloxacin, meropenem R: ampicillin, cotrimoxazole Classification: Definite Gram-negative sepsis	26 year old female **History:** HIV negative, normally fit and well. 2 weeks postpartum. Caesarian section done at QECH. Unwell for 10 days post-operatively: abdominal pain, and fevers. **Examination:** Admission Temperature 40.0°C Recruitment Day 7 Temperature 37.0°C Wound clean. No urinary catheters. Urine dipstick negative. **Bloods:** Day 7 WCC 8.3 x109 L^{-1}, CRP 91 mg L^{-1}, Lactate 2.6mmol L^{-1} **Antibiotics**: 9 days ceftriaxone, 5 days ciprofloxacin **Outcome:** Discharged alive
Participant 2 *E. coli* S: ceftriaxone, chloramphenicol, gentamicin, ciprofloxacin, meropenem R: ampicillin, cotrimoxazole Classification: Definite Gram-negative sepsis	72 year old male **History:** HIV positive, on ART 10 years. Benign prostatic hyperplasia. Unwell for 3-4 weeks: confusion, cough, weight loss, lethargy. Pre-hospital: Co-amoxiclav and azithromycin within 1 month of admission. Week 2 TB Treatment **Examination**: Admission Temperature 35.7°C, GCS 10 Recruitment Day 5 Temperature 37.5°C **Bloods:** Day 5 CD4 158 μL^{-1}, WCC 26.0 x109 L^{-1}, CRP >120mg L^{-1}, lactate 3.4mmol L^{-1}, creatinine 862mmol L^{-1} **Antibiotics:** Ceftriaxone 24 hours **Outcome:** Died in hospital
Participant 3 *E .coli* S: chloramphenicol, gentamicin, amikacin, meropenem, co-amoxiclav R: ceftriaxone, ampicillin, cotrimoxazole, ciprofloxacin Classification: Possible Gram-negative sepsis	32 year old woman **History:** HIV positive, ART 2 years. Mechanical fall into drain, pain in hip and hx of fevers 24 hours later. Associated headache and diarrhoea. Presented to Emergency Department after 5 days, had blood culture and outpatient follow-up arranged. No vital signs available on admission. Discharged on no antibiotics. **Bloods:** Day 6 WCC 6.3 x109 L^{-1}, creatinine 39mmol L^{-1}, CD4 847μL^{-1}, lactate 2.6mmol L^{-1} , CRP 36mg L^{-1}, urine dip negative Repeat blood culture on Day 6 = negative, no antibiotics in between **Antibiotics**: one dose of gentamicin on day 6, then discharged on nothing. **Outcome:** Discharged alive

One approach would be to incorporate uncertainty in true BSI status into a Bayesian latent class model which incorporates the output of the classifications as a latent variable and models the effect on outcome. Another, would be to simply incorporate a classification covariate into logistic regression and survival models and I will consider both approaches in future analyses.

Other explanations for the observation that patients survive without a suitable antibiotic, also need to be considered. I plan to establish MICs for the organisms isolated from participants in the cohort and will relate these to clinical outcomes. Likewise, the organisms have been subject to WGS and I plan a bioinformatic analysis that will look at diversity and clonality of the organisms and their relationship to clinical outcomes.

2. Where should we target interventions aimed at reducing gut mucosal colonisation if 3GC-R in the community setting?

A key finding from my household cohort analysis, was that within-household transmission appears to be a key driver of 3GC-R colonisation and this needs further exploration to determine where interventions aimed at reducing transmission should be targeted. As a first step, the models will need to be refined to include individual level covariates such as antibiotic and CPT use and then fit to the existing data from this study.

As I have discussed in Chapter 7, a diary-based social contact study is needed to challenge the assumption of homogenous mixing used in the models and would provide incredibly useful, and as yet undescribed, social

network information for Blantyre, which could be used in future modelling of not only AMR, but infectious diseases such as TB and SARS-CoV-2.

An important limitation of the modelling I have included in this thesis, is the reliance on phenotypic 3GC-R status, derived from single colony selective culture. Given the highly dynamic process of colonisation observed in this study and elsewhere, it is likely that method will significantly underestimate transmission events by failing to capture the diversity of 3GC-R, both within the household and within the individual. Work from another high prevalence setting, has found a substantial intra-host diversity of *E. coli*, with a median of four sequence types per individual (from 10 colony picks) and different patterns of resistance genes and virulence factors between clones. In my study, I took five colony picks from all individuals and a subset of these have been subject to WGS. I plan an analysis which will describe the diversity of isolates initially within individuals and then households. Ultimately, I aim to use the resolution provided by WGS typing to inform the models and further illuminate where colonisation events are occurring and identify steps in the transmission process which can be interrupted.

References

1. O'Neill J. Review on Antimicrobial Resistance Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. London: Review on Antimicrobial Resistance; 2014. Available from: https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf.

2. Essack SY, Desta AT, Abotsi RE, Agoba EE. Antimicrobial resistance in the WHO African region: current status and roadmap for action. J Public Health (Oxf) **2017**; 39(1): 8-13.

3. Musicha P, Cornick JE, Bar-Zeev N, et al. Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998-2016): a surveillance study. Lancet Infect Dis **2017**; 17(10): 1042-52.

4. Adeolu M, Alnajar S, Naushad S, R SG. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. Int J Syst Evol Microbiol **2016**; 66(12): 5575-99.

5. John E. Bennett RDMJB. Mandell, Douglas, and Bennett's, Principles and Practice of Infectious Diseases: Eighth edition. Philadelphia, PA : Elsevier/Saunders, [2015], **2015**.

6. Day MJ, Hopkins KL, Wareham DW, et al. Extended-spectrum beta-lactamase-producing Escherichia coli in human-derived and foodchain-derived samples from England, Wales, and Scotland: an epidemiological surveillance and typing study. Lancet Infect Dis **2019**; 19(12): 1325-35.

7. Datta N, Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. Nature **1965**; 208(5007): 239-41.

8. Paterson DL, Bonomo RA. Extended-spectrum β-lactamases: A clinical update. Clinical microbiology reviews **2005**; 18(4): 657-86.

9. Kliebe C, Nies BA, Meyer JF, Tolxdorff-Neutzling RM, Wiedemann B. Evolution of plasmid-coded resistance to broad-spectrum cephalosporins. Antimicrobial agents and chemotherapy **1985**; 28(2): 302-7.

10. Knothe H, Shah P, Krcmery V, Antal M, Mitsuhashi S. Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of Klebsiella pneumoniae and Serratia marcescens. Infection **1983**; 11(6): 315-7.

11. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrobial agents and chemotherapy **2010**; 54(3): 969-76.

12. Chow JW, Fine MJ, Shlaes DM, et al. Enterobacter bacteremia: clinical features and emergence of antibiotic resistance during therapy. Annals of internal medicine **1991**; 115(8): 585-90.

13. Meini S, Tascini C, Cei M, Sozio E, Rossolini GM. AmpC beta-lactamase-producing Enterobacterales: what a clinician should know. Infection **2019**; 47(3): 363-75.

14. Wu TL, Siu LK, Su LH, et al. Outer membrane protein change combined with co-existing TEM-1 and SHV-1 beta-lactamases lead to false identification of ESBL-producing Klebsiella pneumoniae. The Journal of antimicrobial chemotherapy **2001**; 47(6): 755-61.

15. Schweizer HP. Understanding efflux in Gram-negative bacteria: opportunities for drug discovery. Expert Opin Drug Discov **2012**; 7(7): 633-42.

16. Lester R, Musicha P, van Ginneken N, et al. Prevalence and outcome of bloodstream infections due to third-generation cephalosporin-resistant Enterobacteriaceae in sub-Saharan Africa: a systematic review. The Journal of antimicrobial chemotherapy **2019**.

17. Heil EL, Johnson JK. Impact of CLSI Breakpoint Changes on Microbiology Laboratories and Antimicrobial Stewardship Programs. Journal of clinical microbiology **2016**; 54(4): 840-4.

18. Adriaenssens N, Coenen S, Tonkin-Crine S, et al. European Surveillance of Antimicrobial Consumption (ESAC): disease-specific quality indicators for outpatient antibiotic prescribing. BMJ Qual Saf **2011**; 20(9): 764-72.

19. Hay SI, Rao PC, Dolecek C, et al. Measuring and mapping the global burden of antimicrobial resistance. BMC Med **2018**; 16(1): 78.

20. Biedenbach DJ, Moet GJ, Jones RN. Occurrence and antimicrobial resistance pattern comparisons among bloodstream infection isolates from the SENTRY Antimicrobial Surveillance Program (1997-2002). Diagnostic microbiology and infectious disease **2004**; 50(1): 59-69.

21. Mammeri H, Van De Loo M, Poirel L, Martinez-Martinez L, Nordmann P. Emergence of plasmid-mediated quinolone resistance in Escherichia coli in Europe. Antimicrobial agents and chemotherapy **2005**; 49(1): 71-6.

22. Paterson DL. Recommendation for treatment of severe infections caused by Enterobacteriaceae producing extended-spectrum beta-lactamases (ESBLs). Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2000**; 6(9): 460-3.

23. Lester R, Haigh K, Wood A, et al. Sustained reduction in third-generation cephalosporin usage in adult inpatients following introduction of an antimicrobial stewardship program in a large urban hospital in Malawi. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2020**.

24. World Health Organization, World Health Organization Model List of Essential Medicines. 2019.

25. Pharmacy Medicines and Poisons Board (Malawi). Restricting use of Second Line ANTI-TB Medicines. 2018 [cited 2019 25th September 2019]; Available from: http://www.pmpb.mw/restricting-use-of-second-line-anti-tb-medicines/.

26. Reddy P, Malczynski M, Obias A, et al. Screening for extended-spectrum beta-lactamase-producing Enterobacteriaceae among high-risk patients and rates of subsequent bacteremia. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2007**; 45(7): 846-52.

27. Denis B, Lafaurie M, Donay JL, et al. Prevalence, risk factors, and impact on clinical outcome of extended-spectrum beta-lactamase-producing Escherichia coli bacteraemia: a five-year study. International journal of

infectious diseases : IJID : official publication of the International Society for Infectious Diseases **2015**; 39: 1-6.

28. Falagas ME, Karageorgopoulos DE. Extended-spectrum beta-lactamase-producing organisms. The Journal of hospital infection **2009**; 73(4): 345-54.

29. Lewis JM, Lester R, Garner P, Feasey NA. Gut mucosal colonisation with extended-spectrum beta-lactamase producing Enterobacteriaceae in sub-Saharan Africa: a systematic review and meta-analysis. Wellcome Open Res **2019**; 4: 160.

30. Albrechtova K, Dolejska M, Cizek A, et al. Dogs of nomadic pastoralists in northern Kenya are reservoirs of plasmid-mediated cephalosporin- and quinolone-resistant Escherichia coli, including pandemic clone B2-O25-ST131. Antimicrobial agents and chemotherapy **2012**; 56(7): 4013-7.

31. Chereau F, Herindrainy P, Garin B, et al. Colonization of extended-spectrum-beta-lactamase- and NDM-1-producing Enterobacteriaceae among pregnant women in the community in a low-income country: a potential reservoir for transmission of multiresistant Enterobacteriaceae to neonates. Antimicrobial agents and chemotherapy **2015**; 59(6): 3652-5.

32. Chirindze LM, Zimba TF, Sekyere JO, et al. Faecal colonization of E. coli and Klebsiella spp. producing extended-spectrum beta-lactamases and plasmid-mediated AmpC in Mozambican university students. BMC infectious diseases **2018**; 18(1): 244.

33. Farra A, Frank T, Tondeur L, et al. High rate of faecal carriage of extended-spectrum beta-lactamase-producing Enterobacteriaceae in healthy children in Bangui, Central African Republic. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2016**; 22(10): 891.e1-.e4.

34. Katakweba AAS, Muhairwa AP, Lupindu AM, et al. First Report on a Randomized Investigation of Antimicrobial Resistance in Fecal Indicator Bacteria from Livestock, Poultry, and Humans in Tanzania. Microbial Drug Resistance **2018**; 24(3): 260-8.

35. Lonchel CM, Meex C, Gangoue-Pieboji J, et al. Proportion of extended-spectrum ss-lactamase-producing Enterobacteriaceae in community setting in Ngaoundere, Cameroon. BMC infectious diseases **2012**; 12: 53.

36. Ouchar Mahamat O, Tidjani A, Lounnas M, et al. Fecal carriage of extended-spectrum beta-lactamase-producing Enterobacteriaceae in hospital and community settings in Chad. Antimicrobial resistance and infection control **2019**; 8: 169.

37. Moremi N, Claus H, Vogel U, Mshana SE. Faecal carriage of CTX-M extended-spectrum beta-lactamase-producing Enterobacteriaceae among street children dwelling in Mwanza city, Tanzania. PloS one **2017**; 12(9): e0184592.

38. Ribeiro TG, Novais Â, Peixe L, Machado E. Atypical epidemiology of CTX-M-15 among Enterobacteriaceae from a high diversity of non-clinical niches in Angola. Journal of Antimicrobial Chemotherapy **2016**; 71(5): 1169-73.

39. Ruppe E, Woerther PL, Diop A, et al. Carriage of CTX-M-15-producing Escherichia coli isolates among children living in a remote village in Senegal. Antimicrobial agents and chemotherapy **2009**; 53(7): 3135-7.

40. Sanneh B, Kebbeh A, Jallow HS, et al. Prevalence and risk factors for faecal carriage of Extended Spectrum beta-lactamase producing Enterobacteriaceae among food handlers in lower basic schools in West Coast Region of The Gambia. PloS one **2018**; 13(8): e0200894.

41. Tellevik MG, Blomberg B, Kommedal O, Maselle SY, Langeland N, Moyo SJ. High Prevalence of Faecal Carriage of ESBL-Producing Enterobacteriaceae among Children in Dar es Salaam, Tanzania. PloS one **2016**; 11(12): e0168024.

42. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E. Fecal Colonization With Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors Among Healthy Individuals: A Systematic Review and Metaanalysis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2016**; 63(3): 310-8.

43. Zhang H, Zhou Y, Guo S, Chang W. High prevalence and risk factors of fecal carriage of CTX-M type extended-spectrum beta-lactamase-producing Enterobacteriaceae from healthy rural residents of Taian, China. Frontiers in microbiology **2015**; 6(MAR).

44. Li B, Sun JY, Liu QZ, Han LZ, Huang XH, Ni YX. High prevalence of CTX-M beta-lactamases in faecal Escherichia coli strains from healthy

humans in Fuzhou, China. Scandinavian journal of infectious diseases **2011**; 43(3): 170-4.

45. Nakayama T, Ueda S, Huong BT, et al. Wide dissemination of extended-spectrum beta-lactamase-producing Escherichia coli in community residents in the Indochinese peninsula. Infection and drug resistance **2015**; 8: 1-5.

46. Ueda S, Ngan BT, Huong BT, Hirai I, Tuyen le D, Yamamoto Y. Limited transmission of bla(CTX-M-9)-type-positive Escherichia coli between humans and poultry in Vietnam. Antimicrobial agents and chemotherapy **2015**; 59(6): 3574-7.

47. McNulty CAM, Lecky DM, Xu-McCrae L, et al. CTX-M ESBL-producing Enterobacteriaceae: estimated prevalence in adults in England in 2014. The Journal of antimicrobial chemotherapy **2018**; 73(5): 1368-88.

48. van Duijkeren E, Wielders CCH, Dierikx CM, et al. Long-term Carriage of Extended-Spectrum beta-Lactamase-Producing Escherichia coli and Klebsiella pneumoniae in the General Population in The Netherlands. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2018**; 66(9): 1368-76.

49. Ny S, Lofmark S, Borjesson S, et al. Community carriage of ESBL-producing Escherichia coli is associated with strains of low pathogenicity: a Swedish nationwide study. The Journal of antimicrobial chemotherapy **2017**; 72(2): 582-8.

50. Babu R, Kumar A, Karim S, et al. Faecal carriage rate of extended-spectrum beta-lactamase-producing Enterobacteriaceae in hospitalised patients and healthy asymptomatic individuals coming for health check-up. Journal of global antimicrobial resistance **2016**; 6: 150-3.

51. Stoesser N, Xayaheuang S, Vongsouvath M, et al. Colonization with Enterobacteriaceae producing ESBLs in children attending pre-school childcare facilities in the Lao People's Democratic Republic. The Journal of antimicrobial chemotherapy **2015**; 70(6): 1893-7.

52. Reuland EA, Al Naiemi N, Kaiser AM, et al. Prevalence and risk factors for carriage of ESBL-producing Enterobacteriaceae in Amsterdam. Journal of Antimicrobial Chemotherapy **2016**; 71(4): 1076-82.

53. Woerther PL, Andremont A, Kantele A. Travel-acquired ESBL-producing Enterobacteriaceae: impact of colonization at individual and community level. J Travel Med **2017**; 24(suppl_1): S29-S34.

54. Kuenzli E, Jaeger VK, Frei R, et al. High colonization rates of extended-spectrum beta-lactamase (ESBL)-producing Escherichia coli in Swiss travellers to South Asia- a prospective observational multicentre cohort study looking at epidemiology, microbiology and risk factors. BMC infectious diseases **2014**; 14: 528.

55. Arcilla MS, van Hattem JM, Haverkate MR, et al. Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. Lancet Infect Dis **2017**; 17(1): 78-85.

56. Huizinga P, van den Bergh MK, van Rijen M, Willemsen I, van 't Veer N, Kluytmans J. Proton Pump Inhibitor Use Is Associated With Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae Rectal Carriage at Hospital Admission: A Cross-Sectional Study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2017**; 64(3): 361-3.

57. Cunningham R, Jones L, Enki DG, Tischhauser R. Proton pump inhibitor use as a risk factor for Enterobacteriaceal infection: a case-control study. The Journal of hospital infection **2018**; 100(1): 60-4.

58. Ben Sallem R, Ben Slama K, Estepa V, et al. Prevalence and characterisation of extended-spectrum beta-lactamase (ESBL)-producing Escherichia coli isolates in healthy volunteers in Tunisia. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology **2012**; 31(7): 1511-6.

59. Zahar JR, Lanternier F, Mechai F, et al. Duration of colonisation by Enterobacteriaceae producing extended-spectrum beta-lactamase and risk factors for persistent faecal carriage. The Journal of hospital infection **2010**; 75(1): 76-8.

60. Lohr IH, Rettedal S, Natas OB, Naseer U, Oymar K, Sundsfjord A. Long-term faecal carriage in infants and intra-household transmission of CTX-M-15-producing Klebsiella pneumoniae following a nosocomial outbreak. The Journal of antimicrobial chemotherapy **2013**; 68(5): 1043-8.

61. Kantele A, Kuenzli E, Dunn SJ, et al. Dynamics of intestinal multidrug-resistant bacteria colonisation contracted by visitors to a high-endemic setting: a prospective, daily, real-time sampling study. Lancet Microbe **2021**; 2(4): e151-e8.

62. Rodriguez-Bano J, Lopez-Cerero L, Navarro MD, Diaz de Alba P, Pascual A. Faecal carriage of extended-spectrum beta-lactamase-producing Escherichia coli: prevalence, risk factors and molecular epidemiology. The Journal of antimicrobial chemotherapy **2008**; 62(5): 1142-9.

63. Valverde A, Grill F, Coque TM, et al. High rate of intestinal colonization with extended-spectrum-beta-lactamase-producing organisms in household contacts of infected community patients. Journal of clinical microbiology **2008**; 46(8): 2796-9.

64. Kurz MS, Bayingana C, Ndoli JM, et al. Intense pre-admission carriage and further acquisition of ESBL-producing Enterobacteriaceae among patients and their caregivers in a tertiary hospital in Rwanda. Tropical medicine & international health : TM & IH **2017**; 22(2): 210-20.

65. Cooper BS, Medley GF, Stone SP, et al. Methicillin-resistant Staphylococcus aureus in hospitals and the community: stealth dynamics and control catastrophes. Proc Natl Acad Sci U S A **2004**; 101(27): 10223-8.

66. Grundmann H, Hellriegel B. Mathematical modelling: a tool for hospital infection control. Lancet Infect Dis **2006**; 6(1): 39-45.

67. Crellen T, Turner P, Pol S, et al. Transmission dynamics and control of multidrug-resistant Klebsiella pneumoniae in neonates in a developing country. Elife **2019**; 8.

68. Bonneault M, Andrianoelina VH, Herindrainy P, et al. Transmission Routes of Extended-Spectrum Beta-Lactamase-Producing Enterobacteriaceae in a Neonatology Ward in Madagascar. The American journal of tropical medicine and hygiene **2019**; 100(6): 1355-62.

69. Haverkate MR, Platteel TN, Fluit AC, et al. Quantifying within-household transmission of extended-spectrum beta-lactamase-producing bacteria. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2017**; 23(1): 46 e1- e7.

70. Mughini-Gras L, Dorado-Garcia A, van Duijkeren E, et al. Attributable sources of community-acquired carriage of Escherichia coli containing beta-lactam antibiotic resistance genes: a population-based modelling study. Lancet Planet Health **2019**; 3(8): e357-e69.

71. de Kraker ME, Stewardson AJ, Harbarth S. Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? PLoS medicine **2016**; 13(11): e1002184.

72. UNSD. Standard Country or Area Codes for Statistics Use. Available from: https://unstats.un.org/unsd/methodology/m49/. **1999**.

73. Acquah SE, Quaye L, Sagoe K, Ziem JB, Bromberger PI, Amponsem AA. Susceptibility of bacterial etiological agents to commonly-used antimicrobial agents in children with sepsis at the Tamale Teaching Hospital. BMC infectious diseases **2013**; 13: 89.

74. Apondi OE, Oduor OC, Gye BK, Kipkoech MK. High Prevalence of Multi-Drug Resistant Klebsiella Pneumoniae in a Tertiary Teaching Hospital in Western Kenya. African journal of infectious diseases **2016**; 10(2): 89-95.

75. Bejon P, Mwangi I, Ngetsa C, et al. Invasive Gram-negative bacilli are frequently resistant to standard antibiotics for children admitted to hospital in Kilifi, Kenya. The Journal of antimicrobial chemotherapy **2005**; 56(1): 232-5.

76. Blomberg B, Manji KP, Urassa WK, et al. Antimicrobial resistance predicts death in Tanzanian children with bloodstream infections: a prospective cohort study. BMC infectious diseases **2007**; 7: 43.

77. Breurec S, Bouchiat C, Sire JM, et al. High third-generation cephalosporin resistant Enterobacteriaceae prevalence rate among neonatal infections in Dakar, Senegal. BMC infectious diseases **2016**; 16(1).

78. Brink A, Moolman GJJ, Cruz da Silva M, et al. Antimicrobial susceptibility profile of selected bacteraemic pathogens from private institutions in South Africa. South African Medical Journal **2007**; 97(4): 273-9.

79. Buys H, Muloiwa R, Bamford C, Eley B. Klebsiella pneumoniae bloodstream infections at a South African children's hospital 2006-2011, a cross-sectional study. BMC infectious diseases **2016**; 16(1): 570.

80. Crichton H, O'Connell N, Rabie H, Whitelaw AC, Dramowski A. Neonatal and paediatric bloodstream infections: Pathogens, antimicrobial resistance patterns and prescribing practice at Khayelitsha District Hospital, Cape Town, South Africa. South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde **2018**; 108(2): 99-104.

81. Dramowski A, Madide A, Bekker A. Neonatal nosocomial bloodstream infections at a referral hospital in a middle-income country: burden, pathogens, antimicrobial resistance and mortality. Paediatrics and international child health **2015**; 35(3): 265-72.

82. Dramowski A, Cotton MF, Rabie H, Whitelaw A. Trends in paediatric bloodstream infections at a South African referral hospital. BMC pediatrics **2015**; 15: 33.

83. Eibach D, Belmar Campos C, Krumkamp R, et al. Extended spectrum beta-lactamase producing Enterobacteriaceae causing bloodstream infections in rural Ghana, 2007-2012. International journal of medical microbiology : IJMM **2016**; 306(4): 249-54.

84. Jaspan HB, Huang LC, Cotton MF, Whitelaw A, Myer L. Bacterial disease and antimicrobial susceptibility patterns in HIV-infected, hospitalized children: A retrospective cohort study. PloS one **2008**; 3(9).

85. Ko WC, Paterson DL, Sagnimeni AJ, et al. Community-acquired Klebsiella pneumoniae bacteremia: global differences in clinical patterns. Emerging infectious diseases **2002**; 8(2): 160-6.

86. Kohli R, Omuse G, Revathi G. Antibacterial susceptibility patterns of blood stream isolates in patients investigated at the Aga Khan University Hospital, Nairobi. East African medical journal **2010**; 87(2): 74-80.

87. Lochan H, Pillay V, Bamford C, Nuttall J, Eley B. Bloodstream infections at a tertiary level paediatric hospital in South Africa. BMC infectious diseases **2017**; 17(1): 750.

88. Marando R, Seni J, Mirambo MM, et al. Predictors of the extended-spectrum-beta lactamases producing Enterobacteriaceae neonatal sepsis at a tertiary hospital, Tanzania. International journal of medical microbiology : IJMM **2018**; 308(7): 803-11.

89. Mhada TV, Fredrick F, Matee MI, Massawe A. Neonatal sepsis at Muhimbili National Hospital, Dar es Salaam, Tanzania; aetiology,

antimicrobial sensitivity pattern and clinical outcome. BMC Public Health **2012**; 12: 904.

90. Morkel G, Bekker A, Marais BJ, Kirsten G, van Wyk J, Dramowski A. Bloodstream infections and antimicrobial resistance patterns in a South African neonatal intensive care unit. Paediatrics and international child health **2014**; 34(2): 108-14.

91. Mshana SE, Kamugisha E, Mirambo M, Chakraborty T, Lyamuya EF. Prevalence of multiresistant gram-negative organisms in a tertiary hospital in Mwanza, Tanzania. BMC research notes **2009**; 2: 49.

92. Ndir A, Diop A, Faye PM, Cisse MF, Ndoye B, Astagneau P. Epidemiology and Burden of Bloodstream Infections Caused by Extended-Spectrum Beta-Lactamase Producing Enterobacteriaceae in a Pediatric Hospital in Senegal. PloS one **2016**; 11(2): e0143729.

93. Obeng-Nkrumah N, Labi AK, Addison NO, Labi JE, Awuah-Mensah G. Trends in paediatric and adult bloodstream infections at a Ghanaian referral hospital: a retrospective study. Annals of clinical microbiology and antimicrobials **2016**; 15(1): 49.

94. Obeng-Nkrumah N, Twum-Danso K, Krogfelt KA, Newman MJ. High levels of extended-spectrum beta-lactamases in a major teaching hospital in Ghana: the need for regular monitoring and evaluation of antibiotic resistance. The American journal of tropical medicine and hygiene **2013**; 89(5): 960-4.

95. Ogunlesi TA, Ogunfowora OB, Osinupebi O, Olanrewaju DM. Changing trends in newborn sepsis in Sagamu, Nigeria: bacterial aetiology, risk factors and antibiotic susceptibility. J Paediatr Child Health **2011**; 47(1-2): 5-11.

96. Onken A, Said AK, Jorstad M, Jenum PA, Blomberg B. Prevalence and Antimicrobial Resistance of Microbes Causing Bloodstream Infections in Unguja, Zanzibar. PloS one **2015**; 10(12): e0145632.

97. Paterson DL, Ko WC, Von Gottberg A, et al. International prospective study of Klebsiella pneumoniae bacteremia: implications of extended-spectrum beta-lactamase production in nosocomial Infections. Annals of internal medicine **2004**; 140(1): 26-32.

98. Perovic O, Singh-Moodley A, Duse A, et al. National sentinel site surveillance for antimicrobial resistance in Klebsiella pneumoniae isolates in South Africa, 2010 - 2012. South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde **2014**; 104(8): 563-8.

99. Preziosi M, Zimba TF, Lee K, et al. A prospective observational study of bacteraemia in adults admitted to an urban Mozambican hospital. South African Medical Journal **2015**; 105(5): 370-4.

100. Sangare SA, Maiga AI, Guindo I, et al. Prevalence of ESBL-producing Enterobacteriaceae isolated from blood cultures in Mali. J Infect Dev Ctries **2016**; 10(10): 1059-64.

101. Seboxa T, Amogne W, Abebe W, et al. High mortality from blood stream infection in Addis Ababa, Ethiopia, is due to antimicrobial resistance. PloS one **2015**; 10(12).

102. Wasihun AG, Wlekidan LN, Gebremariam SA, et al. Bacteriological profile and antimicrobial susceptibility patterns of blood culture isolates among febrile patients in Mekelle Hospital, Northern Ethiopia. SpringerPlus **2015**; 4: 314.

103. Ndir A, Diop A, Ka R, et al. Infections caused by extended-spectrum beta-lactamases producing Enterobacteriaceae: clinical and economic impact in patients hospitalized in 2 teaching hospitals in Dakar, Senegal. Antimicrobial resistance and infection control **2016**; 5: 13.

104. ResistanceMap. Center for Disease Dynamics, Economics & Policy (CDDEP). Available at: http://resistancemap.cddep.org/.

105. Alvarez-Uria G, Gandra S, Laxminarayan R. Poverty and prevalence of antimicrobial resistance in invasive isolates. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases **2016**; 52: 59-61.

106. Goossens H, Ferech M, Vander Stichele R, Elseviers M, Group EP. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet **2005**; 365(9459): 579-87.

107. de Kraker ME, Wolkewitz M, Davey PG, et al. Burden of antimicrobial resistance in European hospitals: excess mortality and length of hospital stay associated with bloodstream infections due to Escherichia coli

resistant to third-generation cephalosporins. The Journal of antimicrobial chemotherapy **2011**; 66(2): 398-407.

108. Stewardson AJ, Allignol A, Beyersmann J, et al. The health and economic burden of bloodstream infections caused by antimicrobial-susceptible and non-susceptible Enterobacteriaceae and Staphylococcus aureus in European hospitals, 2010 and 2011: a multicentre retrospective cohort study. Euro Surveill **2016**; 21(33).

109. Tsai MH, Lee IT, Chu SM, et al. Clinical and Molecular Characteristics of Neonatal Extended-Spectrum beta-Lactamase-Producing Gram-Negative Bacteremia: A 12-Year Case-Control-Control Study of a Referral Center in Taiwan. PloS one **2016**; 11(8): e0159744.

110. Xiao T, Wu Z, Shi Q, et al. A retrospective analysis of risk factors and outcomes in patients with extended-spectrum beta-lactamase-producing Escherichia coli bloodstream infections. Journal of global antimicrobial resistance **2019**; 17: 147-56.

111. Melzer M, Petersen I. Mortality following bacteraemic infection caused by extended spectrum beta-lactamase (ESBL) producing E. coli compared to non-ESBL producing E. coli. The Journal of infection **2007**; 55(3): 254-9.

112. Marra AR, Wey SB, Castelo A, et al. Nosocomial bloodstream infections caused by Klebsiella pneumoniae: impact of extended-spectrum beta-lactamase (ESBL) production on clinical outcome in a hospital with high ESBL prevalence. BMC infectious diseases **2006**; 6: 24.

113. Shrestha P, Cooper BS, Coast J, et al. Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use. Antimicrobial resistance and infection control **2018**; 7: 98.

114. Coast J, Smith RD, Millar MR. Superbugs: should antimicrobial resistance be included as a cost in economic evaluation? Health Econ **1996**; 5(3): 217-26.

115. Maheswaran H, Petrou S, MacPherson P, et al. Cost and quality of life analysis of HIV self-testing and facility-based HIV testing and counselling in Blantyre, Malawi. BMC Med **2016**; 14: 34.

116. Barrasa-Villar JI, Aibar-Remon C, Prieto-Andres P, Mareca-Donate R, Moliner-Lahoz J. Impact on Morbidity, Mortality, and Length of Stay of Hospital-Acquired Infections by Resistant Microorganisms. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2017**; 65(4): 644-52.

117. Maheswaran H, Petrou S, MacPherson P, et al. Economic Costs and Health-Related Quality of Life Outcomes of HIV Treatment After Self- and Facility-Based HIV Testing in a Cluster Randomized Trial. J Acquir Immune Defic Syndr **2017**; 75(3): 280-9.

118. Cosgrove SE, Kaye KS, Eliopoulous GM, Carmeli Y. Health and economic outcomes of the emergence of third-generation cephalosporin resistance in Enterobacter species. Arch Intern Med **2002**; 162(2): 185-90.

119. Graves N, Harbarth S, Beyersmann J, Barnett A, Halton K, Cooper B. Estimating the cost of health care-associated infections: mind your p's and q's. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2010**; 50(7): 1017-21.

120. Gandra S, Barter DM, Laxminarayan R. Economic burden of antibiotic resistance: how much do we really know? Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2014**; 20(10): 973-80.

121. Wozniak TM, Barnsbee L, Lee XJ, Pacella RE. Using the best available data to estimate the cost of antimicrobial resistance: a systematic review. Antimicrobial resistance and infection control **2019**; 8: 26.

122. Barnett AG. Time-dependent exposures and the fixed-cohort bias. Environ Health Perspect **2011**; 119(10): A422-3; author reply A3.

123. Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. J Clin Epidemiol **2008**; 61(12): 1216-21.

124. Barnett AG, Beyersmann J, Allignol A, Rosenthal VD, Graves N, Wolkewitz M. The time-dependent bias and its effect on extra length of stay due to nosocomial infection. Value Health **2011**; 14(2): 381-6.

125. Lambert ML, Suetens C, Savey A, et al. Clinical outcomes of health-care-associated infections and antimicrobial resistance in patients admitted to

European intensive-care units: a cohort study. Lancet Infect Dis **2011**; 11(1): 30-8.

126. De Angelis G, Allignol A, Murthy A, et al. Multistate modelling to estimate the excess length of stay associated with meticillin-resistant Staphylococcus aureus colonisation and infection in surgical patients. The Journal of hospital infection **2011**; 78(2): 86-91.

127. Allegranzi B, Stewardson A, Pittet D. Hand hygiene. Bennett & Brachman's Hospital Infections: Sixth Edition, **2013**.

128. Leistner R, Denkel LA, Gastmeier P, Werner G, Layer F, Pfeifer Y. Prevalence of MRSA and Gram-negative bacteria with ESBLs and carbapenemases in patients from Northern Africa at a German hospital. Journal of Antimicrobial Chemotherapy **2015**; 70(11): 3161-4.

129. Maheswaran H, Petrou S, Cohen D, et al. Economic costs and health-related quality of life outcomes of hospitalised patients with high HIV prevalence: A prospective hospital cohort study in Malawi. PloS one **2018**; 13(3): e0192991.

130. Falagas ME, Kastoris AC, Kapaskelis AM, Karageorgopoulos DE. Fosfomycin for the treatment of multidrug-resistant, including extended-spectrum beta-lactamase producing, Enterobacteriaceae infections: a systematic review. Lancet Infect Dis **2010**; 10(1): 43-50.

131. Stewardson A, Fankhauser C, De Angelis G, et al. Burden of bloodstream infection caused by extended-spectrum beta-lactamase-producing enterobacteriaceae determined using multistate modeling at a Swiss University Hospital and a nationwide predictive model. Infection control and hospital epidemiology **2013**; 34(2): 133-43.

132. Smith RD, Yago M, Millar M, Coast J. Assessing the macroeconomic impact of a healthcare problem: the application of computable general equilibrium analysis to antimicrobial resistance. J Health Econ **2005**; 24(6): 1055-75.

133. Prestinaci F, Pezzotti P, Pantosti A. Antimicrobial resistance: a global multifaceted phenomenon. Pathog Glob Health **2015**; 109(7): 309-18.

134. Luepke KH, Suda KJ, Boucher H, et al. Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications. Pharmacotherapy **2017**; 37(1): 71-84.

135. Coast J, Smith R, Karcher AM, Wilton P, Millar M. Superbugs II: how should economic evaluation be conducted for interventions which aim to contain antimicrobial resistance? Health Econ **2002**; 11(7): 637-47.

136. Drummond MF, Sculpher MJ, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. New York, NY: Oxford University Press; 2015. [Google Scholar].

137. Collaborators GBDCoD. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet **2018**; 392(10159): 1736-88.

138. World Health Organization. AMR Indicators and their relevance to the global indicator framework for the SDGs and targets for the 2030 Agenda for Sustainable Development. 2018. Available at: https://www.who.int/antimicrobial-resistance/interagency-coordination-group/AMR_SDG_indicators_analysis_slides.pdf. Accessed 20 December 2019.

139. Alcohol GBD, Drug Use C. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry **2018**; 5(12): 987-1012.

140. Cassini A, Hogberg LD, Plachouras D, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis **2019**; 19(1): 56-66.

141. Limmathurotsakul D, Dunachie S, Fukuda K, et al. Improving the estimation of the global burden of antimicrobial resistant infections. Lancet Infect Dis **2019**.

142. GLASS method for estimating attributable mortality of antimicrobial resistant bloodstream infections. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.

143. Rottier WC, Ammerlaan HS, Bonten MJ. Effects of confounders and intermediates on the association of bacteraemia caused by extended-spectrum β-lactamase-producing Enterobacteriaceae and patient outcome: a meta-analysis. The Journal of antimicrobial chemotherapy **2012**; 67(6): 1311-20.

144. Turner P, Fox-Lewis A, Shrestha P, et al. Microbiology Investigation Criteria for Reporting Objectively (MICRO): a framework for the reporting and interpretation of clinical microbiology data. BMC Med **2019**; 17(1): 70.

145. Lester R, Maheswaran H, Jewell CP, Lalloo DG, Feasey NA. Estimating the burden of antimicrobial resistance in Malawi: protocol for a prospective observational study of the morbidity, mortality and economic cost of third-generation cephalosporin resistant bloodstream infection. Wellcome Open Res **2020**; 5: 29.

146. Malawi National Statistics Office (n.d). Available at: http://www.nsomalawi.mw Malawi National Statistics Office (n.d). Available at: http://www.nsomalawi.mw

147. UNAIDS. https://www.unaids.org/en/resources/documents/2017/90-90-90. (accessed 27 Dec 2019).

148. UNAIDS. Malawi Country Profile. http://aidsinfo.unaids.org/ (accessed 27 Dec 2019).

149. Chipeta MG, Giorgi E, Mategula D, et al. Geostatistical analysis of Malawi's changing malaria transmission from 2010 to 2017. Wellcome Open Res **2019**; 4: 57.

150. Johnson JG, Kunz LJ, Barron W, Ewing WH. Biochemical differentiation of the Enterobacteriaceae with the aid of lysine-iron-agar. Appl Microbiol **1966**; 14(2): 212-7.

151. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. Med Care **2002**; 40(5): 442-6.

152. MoH, 2018. Malawi Standard Treatment Guidelines https://differentiatedservicedelivery.org/Portals/0/adam/Content/yb4xSSLvE0SW98_z7wTm_w/File/Malawi%20Clinical%20HIV%20Guidelines%202018%20(1).pdf Accessed 14 May 2020.

153. British society for antimicrobial chemotherapy. Methods for Antimicrobial Susceptibility Testing. May 2013. http://bsac.org.uk/wp-content/uploads/2012/02/Version-12-Apr-2013_final.pdf.

154. EUCAST Disk Diffusion Method for Antimicrobial Susceptibility Testing - Version 7.0 (January 2019)

http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Disk_test_documents/2019_manuals/Manual_v_7.0_EUCAST_Disk_Test_2019.pdf. Accessed Dec 28 2019.

155. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters, version 9.0, 2019. http://www.eucast.org/clinical_breakpoints/.

156. Evaluation of a chromogenic medium for extended-spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae. Philippe Lagacé-Wiens et al. University of Manitoba, Canada. ECCMID Poster 2010.

157. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
.

158. de Kraker ME, Wolkewitz M, Davey PG, et al. Clinical impact of antimicrobial resistance in European hospitals: excess mortality and length of hospital stay related to methicillin-resistant Staphylococcus aureus bloodstream infections. Antimicrobial agents and chemotherapy **2011**; 55(4): 1598-605.

159. McIntosh N, Bates P, Brykczynska G, et al. Guidelines for the ethical conduct of medical research involving children. Royal College of Paediatrics, Child Health: Ethics Advisory Committee. Arch Dis Child **2000**; 82(2): 177-82.

160. Bou-Antoun S, Davies J, Guy R, Johnson AP, Sheridan EA, Hope RJ. Descriptive epidemiology of Escherichia coli bacteraemia in England, April 2012 to March 2014. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2016**; 21(35).

161. Marchetti G, Tincati C, Silvestri G. Microbial translocation in the pathogenesis of HIV infection and AIDS. Clinical microbiology reviews **2013**; 26(1): 2-18.

162. Tancrede CH, Andremont AO. Bacterial translocation and gram-negative bacteremia in patients with hematological malignancies. The Journal of infectious diseases **1985**; 152(1): 99-103.

163. Civitarese AM, Ruggieri E, Walz JM, et al. A 10-Year Review of Total Hospital-Onset ICU Bloodstream Infections at an Academic Medical Center. Chest **2017**; 151(5): 1011-7.

164. Fuchs A, Bielicki J, Mathur S, Sharland M, Van Den Anker JN. Reviewing the WHO guidelines for antibiotic use for sepsis in neonates and children. Paediatrics and international child health **2018**; 38(sup1): S3-S15.

165. World Health Organization. (2006). WHO child growth standards : length/height-for-age, weight-for-age, weight-for-length, weight -for-height and body mass index-for-age : methods and development. World Health Organization. https://apps.who.int/iris/handle/10665/43413.

166. Mark Myatt and Ernest Guevarra (2019). zscorer: Child Anthropometry z-Score Calculator. R
package version 0.3.1. https://CRAN.R-project.org/package=zscorer.

167. Benson T, Payongayong E, Ahmed A, Sharma M (2006) Simple household poverty assessment models for Malawi. Conference of the International Association of Agricultural Economists (IAAE). Gold Coast, Australia.

168. National Statistics Office of Malawi (1998) Integrated Household Survey. Zomba, Malawi.

169. Nhlema Simwaka B, Benson T, Salaniponi FM, Theobald SJ, Squire SB, Kemp JR. Developing a socio-economic measure to monitor access to tuberculosis services in urban Lilongwe, Malawi. Int J Tuberc Lung Dis **2007**; 11(1): 65-71.

170. Franco C, Little RJA, Louis TA, Slud EV. Comparative Study of Confidence Intervals for Proportions in Complex Sample Surveys. J Surv Stat Methodol **2019**; 7(3): 334-64.

171. MacPherson P, MacPherson EE, Mwale D, et al. Barriers and facilitators to linkage to ART in primary care: a qualitative study of patients and providers in Blantyre, Malawi. J Int AIDS Soc **2012**; 15(2): 18020.

172. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained
Equations in R. Journal of Statistical Software, 45(3), 1-67. https://www.jstatsoft.org/v45/i03/.

173. Stepf van Buuren KG-O. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software **2011**; 45(3).

174. Marchetti G, Bellistri GM, Borghi E, et al. Microbial translocation is associated with sustained failure in CD4+ T-cell reconstitution in HIV-infected patients on long-term highly active antiretroviral therapy. AIDS **2008**; 22(15): 2035-8.

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

176. Mitchell E, Pearce MS, Roberts A. Gram-negative bloodstream infections and sepsis: risk factors, screening tools and surveillance. Br Med Bull **2019**; 132(1): 5-15.

177. Feasey NA, Houston A, Mukaka M, et al. A reduction in adult blood stream infection and case fatality at a large African hospital following antiretroviral therapy roll-out. PloS one **2014**; 9(3): e92226.

178. Waitt PI, Mukaka M, Goodson P, et al. Sepsis carries a high mortality among hospitalised adults in Malawi in the era of antiretroviral therapy scale-up: a longitudinal cohort study. The Journal of infection **2015**; 70(1): 11-9.

179. Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. The Lancet Infectious Diseases **2010**; 10(6): 417-32.

180. Abernethy JK, Johnson AP, Guy R, Hinton N, Sheridan EA, Hope RJ. Thirty day all-cause mortality in patients with Escherichia coli bacteraemia in England. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2015**; 21(3): 251 e1-8.

181. Tumbarello M, Spanu T, Sanguinetti M, et al. Bloodstream infections caused by extended-spectrum-beta-lactamase-producing Klebsiella pneumoniae: risk factors, molecular epidemiology, and clinical outcome. Antimicrobial agents and chemotherapy **2006**; 50(2): 498-504.

182. Ben-Ami R, Rodríguez-Baño J, Arslan H, et al. A multinational survey of risk factors for infection with extended-spectrum beta-lactamase-producing enterobacteriaceae in nonhospitalized patients. Clinical infectious

diseases : an official publication of the Infectious Diseases Society of America **2009**; 49(5): 682-90.

183. Heijnen M, Cumming O, Peletz R, et al. Shared sanitation versus individual household latrines: a systematic review of health outcomes. PloS one **2014**; 9(4): e93300.

184. S. RP, T.A S, S. M, I.D. A, C.A. B, Sershen. Relationships between shared sanitation facilities and diarrhoeal and soil-transmitted helminth infections: an analytical review. Journal of Water, Sanitation and Hygiene for Development **2019**; 9(2): 198–209.

185. Kim H, Kim YA, Park YS, Choi MH, Lee GI, Lee K. Risk Factors and Molecular Features of Sequence Type (ST) 131 Extended-spectrum beta-Lactamase-producing Escherichia coli in Community-onset Bacteremia. Scientific reports **2017**; 7(1): 14640.

186. Malande OO, Nuttall J, Pillay V, Bamford C, Eley B. A ten-year review of ESBL and non-ESBL Escherichia coli bloodstream infections among children at a tertiary referral hospital in South Africa. PloS one **2019**; 14(9): e0222675.

187. Ng'ambi WF, Ade S, Harries AD, et al. Follow-up and programmatic outcomes of HIV-exposed infants registered in a large HIV centre in Lilongwe, Malawi: 2012-2014. Tropical medicine & international health : TM & IH **2016**; 21(8): 995-1002.

188. Evans C, Jones CE, Prendergast AJ. HIV-exposed, uninfected infants: new global challenges in the era of paediatric HIV elimination. Lancet Infect Dis **2016**; 16(6): e92-e107.

189. Schisterman EF, Perkins NJ, Mumford SL, Ahrens KA, Mitchell EM. Collinearity and Causal Diagrams: A Lesson on the Importance of Model Specification. Epidemiology **2017**; 28(1): 47-53.

190. Altman DG, Andersen PK. Bootstrap investigation of the stability of a Cox regression model. Stat Med **1989**; 8(7): 771-83.

191. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J Epidemiol **2013**; 177(4): 292-8.

192. Therneau T (2020). _A Package for Survival Analysis in R_. R package version 3.1-12, <URL: https://CRAN.R-project.org/package=survival>.

193. Pearl J. An introduction to causal inference. Int J Biostat **2010**; 6(2): Article 7.

194. Johannes Textor and Benito van der Zander (2016). dagitty: Graphical Analysis of Structural
Causal Models. R package version 0.2-2. https://CRAN.R-project.org/package=dagitty.

195. Lewis JM, Henrion M, Rylance J. Validity of the qSOFA Score in Low- and Middle-Income Countries. Jama **2018**; 320(19): 2038-9.

196. Vincent JL, de Mendonca A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med **1998**; 26(11): 1793-800.

197. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Jama **2016**; 315(8): 762-74.

198. Moore CC, Hazard R, Saulters KJ, et al. Derivation and validation of a universal vital assessment (UVA) score: a tool for predicting mortality in adult hospitalised patients in sub-Saharan Africa. BMJ Glob Health **2017**; 2(2): e000344.

199. Bone RC, Balk RA, Cerra FB, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. Chest **1992**; 101(6): 1644-55.

200. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. BMC Med Res Methodol **2008**; 8: 70.

201. Williams TC, Bach CC, Matthiesen NB, Henriksen TB, Gagliardi L. Directed acyclic graphs: a tool for causal studies in paediatrics. Pediatric research **2018**; 84(4): 487-93.

202. Schulz zur Wiesch P, Engelstadter J, Bonhoeffer S. Compensation of fitness costs and reversibility of antibiotic resistance mutations. Antimicrobial agents and chemotherapy **2010**; 54(5): 2085-95.

203. Friedman ND, Temkin E, Carmeli Y. The negative impact of antibiotic resistance. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2016**; 22(5): 416-22.

204. Schwaber MJ, Carmeli Y. Mortality and delay in effective therapy associated with extended-spectrum beta-lactamase production in Enterobacteriaceae bacteraemia: a systematic review and meta-analysis. The Journal of antimicrobial chemotherapy **2007**; 60(5): 913-20.

205. Dautzenberg MJ, Ossewaarde JM, de Kraker ME, et al. Successful control of a hospital-wide outbreak of OXA-48 producing enterobacteriaceae in the Netherlands, 2009 to 2011. Eurosurveillance **2014**; 19(9).

206. de Kraker ME, Davey PG, Grundmann H, group Bs. Mortality and hospital stay associated with resistant Staphylococcus aureus and Escherichia coli bacteremia: estimating the burden of antibiotic resistance in Europe. PLoS medicine **2011**; 8(10): e1001104.

207. Mouton JW, Brown DF, Apfalter P, et al. The role of pharmacokinetics/pharmacodynamics in setting clinical MIC breakpoints: the EUCAST approach. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2012**; 18(3): E37-45.

208. Kristóf K, Szabó D, Marsh JW, et al. Extended-spectrum beta-lactamase-producing Klebsiella spp. in a neonatal intensive care unit: Risk factors for the infection and the dynamics of the molecular epidemiology. European Journal of Clinical Microbiology and Infectious Diseases **2007**; 26(8): 563-70.

209. Fu CM, Tseng WP, Chiang WC, et al. Occult Staphylococcus aureus bacteremia in adult emergency department patients: rare but important. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2012**; 54(11): 1536-44.

210. Dargere S, Cormier H, Verdon R. Contaminants in blood cultures: importance, implications, interpretation and prevention. Clinical

microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases **2018**; 24(9): 964-9.

211. Stewardson AJ, Marimuthu K, Sengupta S, et al. Effect of carbapenem resistance on outcomes of bloodstream infection caused by Enterobacteriaceae in low-income and middle-income countries (PANORAMA): a multinational prospective cohort study. Lancet Infect Dis **2019**; 19(6): 601-10.

212. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis **1987**; 40(5): 373-83.

213. Prin M, Itaye T, Clark S, et al. Critical Care in a Tertiary Hospital in Malawi. World J Surg **2016**; 40(11): 2635-42.

214. Huson MA, Kalkman R, Grobusch MP, van der Poll T. Predictive value of the qSOFA score in patients with suspected infection in a resource limited setting in Gabon. Travel medicine and infectious disease **2017**; 15: 76-7.

215. Huson MAM, Katete C, Chunda L, et al. Application of the qSOFA score to predict mortality in patients with suspected infection in a resource-limited setting in Malawi. Infection **2017**; 45(6): 893-6.

216. Rudd KE, Seymour CW, Aluisio AR, et al. Association of the Quick Sequential (Sepsis-Related) Organ Failure Assessment (qSOFA) Score With Excess Hospital Mortality in Adults With Suspected Infection in Low- and Middle-Income Countries. Jama **2018**; 319(21): 2202-11.

217. Schlapbach LJ, Straney L, Bellomo R, MacLaren G, Pilcher D. Prognostic accuracy of age-adapted SOFA, SIRS, PELOD-2, and qSOFA for in-hospital mortality among children with suspected infection admitted to the intensive care unit. Intensive care medicine **2018**; 44(2): 179-88.

218. van Nassau SC, van Beek RH, Driessen GJ, Hazelzet JA, van Wering HM, Boeddha NP. Translating Sepsis-3 Criteria in Children: Prognostic Accuracy of Age-Adjusted Quick SOFA Score in Children Visiting the Emergency Department With Suspected Bacterial Infection. Front Pediatr **2018**; 6: 266.

219. Management Sciences for Health. International Drug Price Indicator Guide. http://mshpriceguide.org/en/home/. Accessed 14 May 2020.

220. World Bank. Available at: https://www.worldbank.org.

221. Pritchard C, Sculpher M. Productivity Costs: Principles and Practice in Economic Evaluation. London, United Kingdom: Office of Health Economics; 2000. Available at: https://www.ohe.org/publications/productivity-costs-principles-and-practice-economic-evaluation.

222. Group E. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy **1990**; 16(3): 199-208.

223. Jelsma J, Hansen K, De Weerdt W, De Cock P, Kind P. How do Zimbabweans value health states? Popul Health Metr **2003**; 1(1): 11.

224. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ **1996**; 5(2): 141-54.

225. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ **2000**; 320(7243): 1197-200.

226. National Statistics Office of Malawi. Malawi population projections. www.nsomalawi.
mw (accessed 27 Dec 2019).

227. https://data.worldbank.org/topic/poverty.

228. Cassini A, Colzani E, Pini A, et al. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2018**; 23(16).

229. Kostyla C, Bain R, Cronk R, Bartram J. Seasonal variation of fecal contamination in drinking water sources in developing countries: a systematic review. The Science of the total environment **2015**; 514: 333-43.

230. World Health Organisation. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Second edition. Geneva: 2016.

231. Herindrainy P, Rabenandrasana MAN, Andrianirina ZZ, et al. Acquisition of extended spectrum beta-lactamase-producing

enterobacteriaceae in neonates: A community based cohort in Madagascar. PloS one **2018**; 13(3): e0193325.

232. Lonchel CM, Melin P, Gangoue-Pieboji J, Assoumou MC, Boreux R, De Mol P. Extended-spectrum beta-lactamase-producing Enterobacteriaceae in Cameroonian hospitals. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology **2013**; 32(1): 79-87.

233. Schaumburg F, Alabi A, Kokou C, et al. High burden of extended-spectrum beta-lactamase-producing Enterobacteriaceae in Gabon. The Journal of antimicrobial chemotherapy **2013**; 68(9): 2140-3.

234. Mushi MF, Mwalutende AE, Gilyoma JM, et al. Predictors of disease complications and treatment outcome among patients with chronic suppurative otitis media attending a tertiary hospital, Mwanza Tanzania. BMC ear, nose, and throat disorders **2016**; 16: 1.

235. Saito R, Koyano S, Nagai R, Okamura N, Moriya K, Koike K. Evaluation of a chromogenic agar medium for the detection of extended-spectrum ss-lactamase-producing Enterobacteriaceae. Letters in applied microbiology **2010**; 51(6): 704-6.

236. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press 2007.

237. Becker N. A general chain binomial model for infectious diseases. Biometrics **1981**; 37(2): 251-8.

238. Keeling MJ, Eames KT. Networks and epidemic models. J R Soc Interface **2005**; 2(4): 295-307.

239. Keeling M. The implications of network structure for epidemic dynamics. Theor Popul Biol **2005**; 67(1): 1-8.

240. Pearl J. Causal inference from indirect experiments. Artif Intell Med **1995**; 7(6): 561-82.

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

242. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings DA. Close encounters of the infectious kind: methods to measure social mixing behaviour. Epidemiology and infection **2012**; 140(12): 2117-30.

243. Eames KT, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured dynamic social contact patterns explain the spread of H1N1v influenza. PLoS Comput Biol **2012**; 8(3): e1002425.

244. Edmunds WJ, O'Callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proc Biol Sci **1997**; 264(1384): 949-57.

245. Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS medicine **2008**; 5(3): e74.

246. Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Soc Sci Med **1985**; 21(11): 1203-16.

247. Turner P, Ashley EA, Celhay OJ, et al. ACORN (A Clinically-Oriented Antimicrobial Resistance Surveillance Network): a pilot protocol for case based antimicrobial resistance surveillance. Wellcome Open Res **2020**; 5: 13.