Identifying and characterising sexual transmission of enteric pathogens in men who have sex with men using classical and molecular epidemiological methods

Holly Diana Mitchell

Thesis submitted for the degree of Doctor of Philosophy

UCL

2020

Declaration of authorship

I, Holly Diana Mitchell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Date:

Abstract

Enteric pathogens are transmitted via the faecal-oral route and commonly cause diarrhoea and/or vomiting. In recent years, there have been numerous outbreaks in men who have sex with men (MSM), primarily *Shigella* spp., often associated with antimicrobial resistance. My research aimed to investigate the characteristics, risk factors and burden of bacterial enteric pathogens (BEPs) in MSM that could inform control.

I conducted a cross-sectional study at a London sexual health clinic (SHC) to estimate the prevalence of BEPs in MSM and associated risk factors. I linked whole genome sequencing (WGS) data with clinical and behavioural data on reported cases of *Shigella flexneri* to i) characterise transmission within sexual networks of MSM and ii) validate a public health tool for identifying MSM clusters in near real-time.

One in 10 predominantly asymptomatic MSM attending the SHC had a BEP detected, which was associated with higher-risk sexual behaviours. Among MSM with a BEP, presence of a genotypic marker of azithromycin resistance was associated with a history of bacterial sexually transmitted infections (STIs). In the WGS study, *S. flexneri* isolates from MSM largely belonged to two clades associated with genotypic markers of azithromycin resistance, with evidence of sustained transmission through sex between men. Over one third of isolates within MSM clades were from people living with HIV. The public health tool distinguished MSM from non-sexual transmission clusters.

My research provides strong evidence that BEPs are transmitted sexually in MSM and asymptomatic carriage may be sustaining transmission. The findings suggest that antimicrobial treatment for STIs selected for resistance in gut organisms, emphasising the need for better antimicrobial stewardship. Phylogenetic analyses provided novel insights about *S. flexneri* transmission in sexual networks of MSM that could inform clinical care and public health management. Real-time identification of MSM clusters might inform the delivery of rapid and appropriate responses.

Impact statement

Following high profile and sustained outbreaks of bacterial enteric pathogens (BEPs), especially *Shigella* spp., among men who have sex with men (MSM), I conducted a cross-sectional study at the UK's largest sexual health clinic to estimate the prevalence of BEPs and associated risk factors. My study provides the most robust estimates of BEP prevalence among MSM in England to date and contributes towards our understanding of their spread and persistence in this population. One in 10 MSM had a BEP detected and most did not report gastrointestinal symptoms, consistent with the theory that asymptomatic carriage might play a key role in sustaining transmission in MSM. Furthermore, BEP detection was associated with a suite of higher-risk sexual behaviours, providing evidence that these pathogens are being transmitted through sexual contact. Identifying MSM who are at higher risk of BEPs might allow better targeting of interventions that aim to control transmission.

In England, whole genome sequencing (WGS) is routinely performed for the public health surveillance of *Shigella* spp. For the first time, I combined WGS data for all *Shigella flexneri* isolates referred to Public Health England (PHE) over a two-year period with rich epidemiological data, including information on sexual behaviour and HIV infection. My analyses provide unique population-level insights into the molecular, clinical and epidemiological characteristics associated with different transmission networks that could be used to inform patient management and the delivery of a more effective and targeted public health response.

My findings will inform best practice for *Shigella* spp. cluster investigations in England and will be relevant to any country developing and implementing WGS for public health surveillance and outbreak detection. I used robust data on sexual identity and behaviour from reported *S. flexneri* cases to validate a WGS tool for detecting and distinguishing MSM clusters from other non-sexual transmission clusters. This approach is suitable for use by PHE to inform the delivery of a more timely and appropriate public health response. In addition, the tool could be used to improve the targeting of appropriate health care management and advice given to cases, including men who do not identify as gay or bisexual.

My research has improved our understanding about the development and spread of antimicrobial resistance in MSM and emphasises the need for better antimicrobial stewardship in this population. My cross-sectional study explored the prevalence of *mphA,* a genotypic marker of azithromycin resistance, in a large population of predominantly asymptomatic MSM, and how this relates to BEP detection and a history of sexually transmitted infections (STIs). Among MSM with a BEP, *mphA* was strongly associated with a bacterial STI diagnosis in the past year, which might reflect previous antimicrobial exposure acting as a selective pressure on gut organisms. My WGS analyses found that azithromycin resistance was strongly associated with *S. flexneri* isolates belonging to sexual networks of MSM. My findings emphasise the need to take a holistic approach that considers the long-term consequences of frequent antimicrobial exposure in high-risk populations, specifically the development of resistance in both target and non-target pathogens.

Table of contents

Declaration of authorship 2

Abstract 3

Impact statement 4

Table of contents 6

List of tables 9

List of figures 11

List of boxes 12

List of abbreviations 13

Publications and conference presentations arising from this research 15

Funding 16

Acknowledgements 17

Chapter 1: Introduction 19

1.1 Background and rationale 19

1.2 Research questions 22

1.3 Aim and objectives 23

1.4 Structure of the thesis 23

1.5 My role in this research 26

Chapter 2: Background 28

2.1 Introduction to enteric pathogens 28

2.2 Aetiology and clinical features 29

2.3 Diagnosis and clinical management 33

2.4 Detection and surveillance of enteric pathogens in England 34

2.5 Incidence of enteric infections in the UK 40

2.6 Prevention and control 44

2.7 Distribution and characteristics of enteric pathogens in MSM 48

Chapter 3: Prevalence and risk factors of bacterial enteric pathogens in MSM: a cross-sectional study at a London sexual health clinic 74

3.1 Introduction and rationale 74

3.2 Aim and objectives 75

3.3 Methods 76

3.4 Results 95

3.5 Discussion 121

Chapter 4: Use of molecular epidemiology to understand the distribution and genetic diversity of *Shigella flexneri* in MSM 136

4.1 Introduction 136

4.2 Study rationale 143

4.3 Aim and Objectives 144

4.4 Methods 145

4.5 Results 151

4.6 Discussion 174

Chapter 5: Validation of a SNP clustering algorithm for identifying sexual transmission of *Shigella flexneri* in MSM 184

5.1 Introduction and rationale 184

5.2 Aim and Objectives 186

5.3 Methods 187

5.4 Results 189

5.5 Discussion 198

Chapter 6: Discussion 204

6.1 Summary of PhD research 204

6.2 Clinical and public health implications 205

6.3 Further research 217

6.4 Overall conclusions 220

References 221

Appendices 250

Appendix for Chapter 2 250

Appendix for Chapter 3 255

Appendix for Chapter 4 293

List of tables

Table 2.1: Selected enteric pathogens of public health importance in England 31

Table 2.2: Data sources for the national surveillance of enteric pathogens in England 35

Table 2.3: Laboratory confirmed cases of selected enteric pathogens in England, 2017 39

Table 2.4: Risk groups for the transmission of enteric pathogens 47

Table 2.5: Aetiological agents of sexually transmissible enteric infections 50

Table 2.6: Excess number of male cases, male-to-female ratio and percentage male by pathogen, risk area and age group, for laboratory-confirmed gastrointestinal infections with no reported travel history, England, 2003-2013 68

Table 3.1: GUMCAD surveillance system data items 83

Table 3.2: Clinical and behavioural data items extracted from the clinic database at Dean Street 84

Table 3.3: Proportion of missing data for behavioural and clinical variables with more than 5% missing data 97

Table 3.4: Socio-demographic, clinical and behavioural characteristics of the study population 99

Table 3.5: Comparison of socio-demographic characteristics of the study population with all eligible MSM clinic attendees during the study period 103

Table 3.6: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen 108

Table 3.7: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in HIV-negative/unknown status MSM 113

Table 3.8: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in MSM living with HIV 118

Table 3.9: Detection of bacterial enteric pathogens in the Dean Street study and in participants aged 15 to 74 years in the IID1 study 128

Table 3.10: Detection of bacterial enteric pathogens in the Dean Street study and in participants aged five years and over in the IID2 study 129

Table 3.11: Bacterial enteric pathogens detected in the Dean Street study and among MSM attending a Melbourne sexual health clinic 130

Table 4.1: Summary of main terms and their definitions 148

Table 4.2: Epidemiological characteristics of *S. flexneri* CC245 isolates with and without linked questionnaire data in England, August 2015 to July 2017.. 153

Table 4.3: Selected epidemiological, clinical and molecular characteristics for cases with questionnaire data... 158

Table 4.4: Epidemiological and molecular characteristics of people living with HIV 160

Table 4.5: Epidemiological and molecular characteristics of 10-SNP single linkage clusters± nested within travel-associated lineages and containing isolates from confirmed MSM 162

Table 4.6: Clinical characteristics, antimicrobial treatment and health-seeking behaviour among *S. flexneri* cases with questionnaire data.. 166

Table 4.7: Characteristics associated with blood in stools, mucus in stools and fever in adults diagnosed with *S. flexneri* and with linked questionnaire data 169

Table 4.8: Characteristics associated with hospital admission and antimicrobial use in adults diagnosed with *S. flexneri* and with linked questionnaire data 171

Table 5.1: Size and distribution of clusters by SNP threshold cut-off and cluster classification 191

Table 5.2: Sexual identity, recent sexual contact and foreign travel for clustered cases with a completed questionnaire 194

Table 5.3: Impact of SNP threshold cut-off on the size, distribution and classification of the cluster 195

List of figures

Figure 1.1: Schematic representation of the framework for this thesis 25

Figure 2.1: Notification of Infectious Diseases (NOIDS) by registered medical practitioners 37

Figure 2.2: The surveillance pyramid 40

Figure 2.3: Non-travel associated diagnoses of *S. flexneri* serotype 2a, *S. flexneri* serotype 3a and *S. sonnei* in England (≥16 years old), by year, 2004-2018 65

Figure 3.1: Data management, linkage and anonymisation algorithm for rectal swab specimens and epidemiological data collected at Dean Street 86

Figure 3.2: Flow chart showing the number of specimens processed, assessed for eligibility and PCR tested 95

Figure 3.3: Detection of bacterial enteric pathogens by clinic attended 104

Figure 3.4: Detection of *mphA* by bacterial STI history in the past year stratified by bacterial enteric pathogen detection 121

Figure 4.1: Illustration of the four steps in the Illumina whole genome sequencing process 138

Figure 4.2: Total number of *S. flexneri* isolates and the number with linked questionnaire data 152

Figure 4.3: Phylogeny of *S. flexneri* CC245 isolates, August 2015 to July 2017 157

Figure 4.4: Detection of novel strain transmission among MSM 163

Figure 4.5: Timed phylogeny estimating the introduction of a novel lineage into a presumed MSM network 164

Figure 5.1: Illustration of the SNP address 185

Figure 5.2: Sensitivity and specificity of the cluster classification tool at different SNP thresholds using the questionnaire data as the gold standard 193

Figure 5.3: Phylogeny of *S. flexneri* CC245 showing the distribution of clusters, genotypic markers of azithromycin resistance and gay-identifying men 196

Figure 5.4: Phylogeny of 25-SNP ‘adult male’ cluster 34.42.42.42.% 197

List of boxes

Box 2.1: Hygiene advice to prevent the person-to-person spread of enteric pathogens....... 46

Box 4.1: Examples of the use WGS for improving case ascertainment and identifying linked cases during outbreak investigations.. 142

List of abbreviations

AIDS	Acquired immunodeficiency syndrome
AMR	Antimicrobial resistance
BASHH	British Association for Sexual Health and HIV
BEP	Bacterial enteric pathogen
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CRN	Clinical Research Network
DNA	Deoxyribonucleic acid
DS	Dean Street
DSE	Dean Street Express
EAEC	Enteroaggregative *E. coli*
ECDC	European Centre for Disease Prevention and Control
EHEC	Enterohaemorrhagic *E. coli*
EHO	Environmental health officer
EIA	Enzyme immunoassay
EPEC	Enteropathogenic *E. coli*
GBRU	Gastrointestinal Bacteria Reference Unit
GBS	Guillain-Barré syndrome
GDW	Gastro Data Warehouse
GP	General practice
HAART	Highly active antiretroviral therapy
HARS	HIV and AIDS Reporting System
HAV	Hepatitis A virus
HIV	Human immunodeficiency virus
HPRU	Health Protection Research Unit
HPT	Health Protection Team
HRA	Health Research Authority
HUS	Haemolytic uraemic syndrome
IID	Infectious Intestinal Disease
IID1	The first study of Infectious intestinal disease in the community
IID2	The second study of infectious intestinal disease in the community
IMD	Index of Multiple Deprivation
IQR	Inter-quartile range

LA	Local Authority
LGV	Lymphogranuloma venereum
LSOA	Lower Layer Super Output Area
MLST	Multi-locus sequence typing
MSM	Men who have sex with men
NATSAL	National Survey of Sexual Attitudes and Lifestyle
NHS	National Health Service
NIHR	National Institute for Health Research
NOIDs	Notifications of Infectious Diseases
NWLP	North West London Pathology
ONS	Office for National Statistics
OR (aOR)	Odds ratio (adjusted odds ratio)
PCR	Polymerase Chain Reaction
PHE	Public Health England
PR (aPR)	Prevalence ratio (adjusted prevalence ratio)
PrEP	Pre-exposure prophylaxis
R&D	Research and development
R	Case reproduction number
R_0	Basic reproduction number
REC	Research ethics committee
RMP	Registered medical practitioner
SGSS	Second Generation Surveillance System
SHC	Sexual Health Clinic
SHHAPT	Sexual Health and HIV Activity Property Type Code
SNP	Single nucleotide polymorphism
SOP	Standard operating procedure
STEC	Shiga toxin-producing *E. coli*
STI	Sexually transmitted infection
THT	Terrence Higgins Trust
UCL	University College London
WGS	Whole genome sequencing
WHO	World Health Organisation

Publications and conference presentations arising from this research

The following manuscripts have been published in a scientific peer-reviewed journal:

1. **Mitchell H**, Hughes G. Recent epidemiology of sexually transmissible enteric infections in men who have sex with men. *Curr Opin Infect Dis*. 2018;31(1):50-6.

2. **Mitchell HD**, Mikhail AFW, Painset A, Dallman TJ, Jenkins C, Thomson NR, Field N, Hughes, G. Use of whole genome sequencing to identify clusters of *Shigella flexneri* associated with sexual transmission in men who have sex with men in England: a validation study using linked behavioural data. *Microbial Genomics* 2019;5(11).

The following abstracts were accepted for oral presentations at the STI & HIV 2019 World Congress - Joint Meeting of the 23rd International Society for Sexually Transmitted Diseases Research (ISSTDR) & 20th International Union against Sexually Transmitted Infections (IUSTI), July 14-17, Vancouver, Canada:

1. **Mitchell HD**, Whitlock G, Zdravkov J, Olsson J, Jenkins C, Thomson NR, Field N, Hughes G. Prevalence of sexually transmissible enteric infections in men who have sex with men (MSM): Preliminary findings from a cross-sectional study.

2. **Mitchell HD**, Mikhail AFW, Painset A, Dallman TJ, Jenkins C, Thomson NR, Field N, Hughes G. Use of whole genome sequencing to identify clusters of *Shigella flexneri* associated with sexual transmission in men who have sex with men.

Funding

My PhD was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Blood Borne and Sexually Transmitted Infections at University College London (UCL) in partnership with Public Health England (PHE) and in collaboration with the London School of Hygiene and Tropical Medicine. All laboratory work including shipping and testing of biological specimens was funded by the NIHR HPRU in Gastrointestinal Infections at University of Liverpool in partnership with PHE, in collaboration with University of East Anglia, University of Oxford and the Quadram Institute. The views expressed in this thesis are my own and not necessarily those of the National Health Service (NHS), the NIHR, the Department of Health or PHE.

Acknowledgements

I would like to thank the many people who have supported me throughout this PhD research.

Thank you to my supervisors Gwenda Hughes, Nigel Field, Claire Jenkins and Nick Thomson for their guidance, advice and encouragement. I have benefitted immensely from their combined expertise and have learnt so much from them.

I would like to thank the NIHR HPRUs in Blood Borne and Sexually Transmitted Infections and Gastrointestinal Infections for funding my PhD. I am also grateful to the UCL Institute for Global Health Student Conference Fund & School of Life and Medical Sciences Graduate Conference Fund for funding my journey to Vancouver to present my research.

At PHE, thank you to Tim Dallman and Anaïs Painset for welcoming me into the bioinformatics team and for teaching me everything I know about phylogenetics. I am also grateful to the GBRU team who were incredibly supportive during my time in lab, particularly Panida Silalang who helped me to master the art of bacterial DNA extraction and PCR testing and who provided many enjoyable lunches (outside of the lab!), and Hafsah Begum who helped me to print thousands of specimen labels. Thank you also to Megan Bardsley, Paula Blomquist and Paul Vanta for providing GUMCAD data, Stephen Duffell for facilitating seamless data transfer, Amy Mikhail for helping with the whole genome sequencing cluster analysis and Peter Kirwan for facilitating HIV data linkage. Thank you to Hamish Mohammed for his support and encouragement throughout this PhD process and for providing tips on the best coffee spots.

Thank you so much to the team at Dean Street for collaborating with me on my study. This would not have been possible without the efforts of Gary Whitlock. Thank you to Gary, Jey Zdravkov and Jenny Olsson for dedicating many hours to data extraction. I would also like to thank those who helped to get the study up and running: Miriam Curran, Rebecca Harris, Claire McGhee and Michael Post.

Thank you to Chelsea and Westminster NHS Research and Development for supporting the Dean Street study, especially Alexandra Schoolmeesters and Serah Duro for all their hard work extracting data. I am grateful to Lorraine Omari-Asor and the NIHR Clinical Research

Network for supporting the study. Thank you also to Monica Rebec and Nigel Bartholomew at North West London Pathology for sending me specimens.

At UCL, thank you to Cath Mercer for advice with data analysis and for providing Natsal-3 data, Andrew Copas for providing statistical advice, and to my fellow PhD students and colleagues for providing support, motivation and laughter, particularly Melissa Cabecinha, Sara Croxford, Tyrone Curtis, Seth Francis-Graham and Katy Town.

To my friends and family, thanks for listening and always putting things into perspective. You have provided much needed fun, humour and distraction away from this PhD. Special thanks go to Caroline and Emma for their top proof reading skills. To Alan, I would not have been able to do this PhD without your support. Thank you for putting up with my constant worrying and for doing all the cooking - it is time for a holiday.

Chapter 1: Introduction

1.1 Background and rationale

Enteric pathogens are transmitted via the faecal-oral route and commonly cause diarrhoea and/or vomiting. In high-income countries, many cases are linked to foreign travel to regions considered to be at high-risk of enteric disease, primarily South Asia or sub-Saharan Africa.[1,2] However, these pathogens can also be transmitted through sexual contact, which occurs through the ingestion of faecal matter directly linked to sexual activity. Men who have sex with men (MSM) are particularly at risk because they may engage in sexual practices that facilitate more direct faecal-oral transmission. This concept was first described in the 1970s,[3] however, within the last two decades there have been an increasing number of enteric pathogen outbreaks among MSM worldwide, including *Shigella* spp.,[4-7] hepatitis A virus,[8-10] *Campylobacter* spp.,[11,12] Shiga toxin-producing *Escherichia coli* (STEC)[13] and *Entamoeba histolytica.*[14]

Our current understanding of the epidemiology of enteric infections in MSM is primarily based on analyses of laboratory surveillance data, clinical case reports and information collected during public health follow-up of outbreaks, particularly for shigellosis (the term used to describe diarrhoeal illness caused by infection with *Shigella* spp. bacteria [see section 2.2]). In England, a national outbreak of domestically acquired *Shigella flexneri* serotype 3a, which started in 2009, was later shown to be associated with sexual transmission among MSM.[15] Since 2011, increasing laboratory reports of *S. flexneri* serotype 2a and *S. sonnei* in adult men with no foreign travel history suggest an intensification of the shigellosis epidemic in MSM through separate introductions to the population.[5] Furthermore, a small cluster of STEC O117:H7 serotype during 2013 to 2014 was associated with sexual transmission among MSM.[13] The behavioural profile of cases in these outbreaks was similar, including specific sexual practices and drug-use behaviours predominantly among MSM living with human immunodeficiency virus (HIV).[4,13] More recently, a large international outbreak of hepatitis A, predominantly associated with sexual transmission among MSM, was reported across Europe (including the UK) and beyond including regions of Chile, North America, Australia, Israel and

Taiwan.[16-19] These outbreaks have occurred alongside epidemics of other sexually transmitted infections (STIs) in MSM including gonorrhoea, syphilis and lymphogranuloma venereum (LGV), consistent with overlapping sexual networks.[20,21]

As well as the growing number of reported outbreaks, the emergence and spread of resistance to first- or second-line antimicrobials, including fluoroquinolones and macrolides in some pathogens, is a public health concern (e.g. *Shigella* spp. and *Campylobacter* spp.).[11-13,22,23] Of note is the development of resistance to antimicrobials that are not recommended first-line treatments, such as resistance to the macrolide azithromycin in MSM-associated strains of *Shigella* spp.[22,24] Shigellosis is often managed conservatively but where clinically indicated, the primary treatment is ciprofloxacin (a fluoroquinolone). On the other hand, azithromycin is used for the treatment of several bacterial STIs. Antimicrobial treatment for one bacterial infection may have implications for the development of antimicrobial resistance (AMR) in other infections by creating an environment that selects for resistance. Genotypic markers of AMR can then be transferred to other bacteria and spread further.[24] As such, it is hypothesised that azithromycin resistance in MSM-associated *Shigella* spp. was selected through off-target antimicrobial treatment for bacterial STIs.

In England, enteric pathogens are statutorily notifiable to enable prompt and appropriate public health action.[25] Despite this, it is unclear whether sexual transmission is routinely considered as a potential route of infection, either by patients or by healthcare professionals during case follow-up investigations. Inadequate consideration of sexual transmission was demonstrated during the investigation of the 2009 *S. flexneri* serotype 3a outbreak, where most general practitioners (GPs) assumed that the men had food poisoning.[4] This is likely to lead to missed or mistaken diagnoses, and sub-optimal clinical management, health promotion advice and contact tracing. Identifying and distinguishing probable sexual acquisition of infection from other types of faecal-oral transmission (e.g. food-borne) is essential to ensure that people receive appropriate advice on preventing onward transmission and to inform subsequent public health responses. In addition, MSM diagnosed with an enteric pathogen are likely to be at risk of other STIs and HIV, and the clinical outcomes and management of enteric pathogens may be affected by HIV.[26,27] Identification of sexual exposure would therefore facilitate

appropriate referral to sexual health clinics (SHCs) for further testing, partner notification and appropriate management.

Whilst surveillance data have provided us with a crude understanding of the burden of enteric pathogens in MSM, predominantly *Shigella* spp., there remain a number of unanswered questions about the underlying prevalence and transmission dynamics of enteric pathogens in MSM. This is because 1) routine surveillance data only represent symptomatic individuals who present to healthcare and have a stool sample collected for microbiological investigations, and where an infectious pathogen is detected (culture is not always successful), 2) sexual identity and behaviour are not routinely collected in follow-up investigations, which means that sexual transmission must be inferred at a population level using the gender ratio and the excess number of adult male cases,[5] and 3) there is currently no routine screening for asymptomatic carriage, not least because the clinical implications and risk of onward transmission are not well understood.

A previous UK-based feasibility study estimated the prevalence of bacterial enteric pathogens (BEPs) in a convenience sample of MSM diagnosed with rectal chlamydial infection at selected SHCs in 2012.[28] Of 444 residual rectal swabs tested using real-time Polymerase Chain Reaction (PCR), the overall prevalence of BEPs was 8.6% (95% CI: 6.3% to 11.6%): 1.8% (95% CI: 0.9%-3.6%) for *Shigella* spp., 1.8% (95% CI: 0.9% to 3.6%) for *Campylobacter* spp., and 5.2% (95% CI: 3.5% to 7.7%) for enteroaggregative *E. coli* (EAEC). About half of the specimens that had a pathogen detected were from cases that did not report symptoms suggesting that asymptomatic carriage may play a role in sustaining transmission among MSM.[28] The study demonstrated that rectal swabs could be used to detect BEPs, however, use of a small, biased convenience sample with no behavioural information did not enable a comprehensive or representative assessment of the burden and risk factors associated with BEPs among MSM in the UK. Furthermore, while outbreak investigations and analyses of surveillance data have described specific sexual risk practices and contextual factors of MSM diagnosed with enteric pathogens in the UK, these have been without a suitable comparison group of uninfected individuals and it has not been possible to perform unbiased risk factor analyses to inform infection control efforts.

Given the importance of controlling AMR, there is also a need to better understand how the acquisition of genetic elements that encode AMR determinants in BEPs relates to prior STI infection and history of antimicrobial exposure to inform clinical prescribing practice and foster good antimicrobial stewardship.

From a public health perspective, insight into the prevalence of enteric pathogens in MSM, and the individual behavioural and clinical risk factors and network characteristics associated with sexual transmission, could help inform the design, development and delivery of appropriately tailored clinical and public health interventions that seek to control enteric pathogens and might inform guidelines on antimicrobial treatment. BEPs are a subset of all enteric pathogens that have been associated with high profile and sustained outbreaks among MSM, particularly successive epidemics of different *S. flexneri* serotypes and *S. sonnei*. Given this, and the time constraints associated with completing a PhD, my research focussed on BEPs, and included two studies on *S. flexneri*.

1.2 Research questions

The following research questions underpinned the research presented in this thesis:

1. What is the overall prevalence of BEPs in MSM?
2. What is the role of asymptomatic carriage or subclinical infection in sustaining transmission of BEPs among MSM?
3. What is the relationship between BEPs and sexual risk behaviours in MSM and how does this overlap with STIs and HIV?
4. What is the relationship between azithromycin resistance in BEPs in MSM and previous treatment for STIs?
5. What molecular, clinical and epidemiological characteristics of reported cases are indicative of transmission in sexual networks of MSM?
6. How can public health tools utilising whole genome sequencing (WGS) data be used to identify and discriminate sexual from non-sexual transmission of BEPs?

1.3 Aim and objectives

The aim of this PhD research was to investigate and describe the clinical and epidemiological characteristics, risk factors and burden of infection associated with BEPs in MSM that could inform the development, targeting and delivery of more appropriate and effective interventions.

The objectives were to:

1. Provide robust and representative estimates of prevalence to assess the burden of BEPs in MSM
2. Investigate the potential for subclinical infection or asymptomatic carriage in sustaining transmission of BEPs in MSM
3. Describe the clinical, behavioural and epidemiological risk factors of BEPs in MSM
4. Explore the association between BEPs in MSM and genotypic markers of antimicrobial resistance, and how this relates to a previous bacterial STI diagnosis
5. Improve identification and characterisation of sexual versus other types of transmission of BEPs

1.4 Structure of the thesis

My PhD research consisted of a cross-sectional study of BEPs among MSM attending a large SHC in central London, and novel analyses of WGS data of *S. flexneri* isolates referred to the PHE national reference laboratory and sequenced as part of national surveillance. Figure 1.1 provides a schematic representation of how these two components relate to my research questions and objectives. The thesis is structured as follows:

Chapter 2 sets out the context and rationale for my thesis. It includes an overview of enteric pathogens in England, and a comprehensive review of the literature on the epidemiology of enteric pathogens in MSM, including the distribution and characteristics associated with recent outbreaks and the key issues for prevention and control.

Chapter 3 describes the design, implementation and analysis of a cross-sectional study carried out in MSM attending a large sexual health and HIV service in central London (Dean

Street Sexual Health and HIV Service, part of Chelsea and Westminster Hospital NHS Foundation Trust) for routine sexual health check-ups. The aim of this study was to better understand the overall burden of BEPs in sexually active MSM and the associated risk factors.

Chapter 4 presents phylogenetic and epidemiological analyses of WGS data of *S. flexneri* isolates and linked case questionnaires to characterise sexual networks of MSM and describes how these differ from other types of transmission.

Chapter 5 utilises *S. flexneri* WGS data and linked case questionnaires (also used in Chapter 4) to validate a real-time public health tool for discriminating clusters of cases linked through sexual and non-sexual transmission to inform a rapid public health response.

Chapter 6 discusses the overall findings from the PhD research within the context of the wider clinical and public health implications, and highlights opportunities for further research.

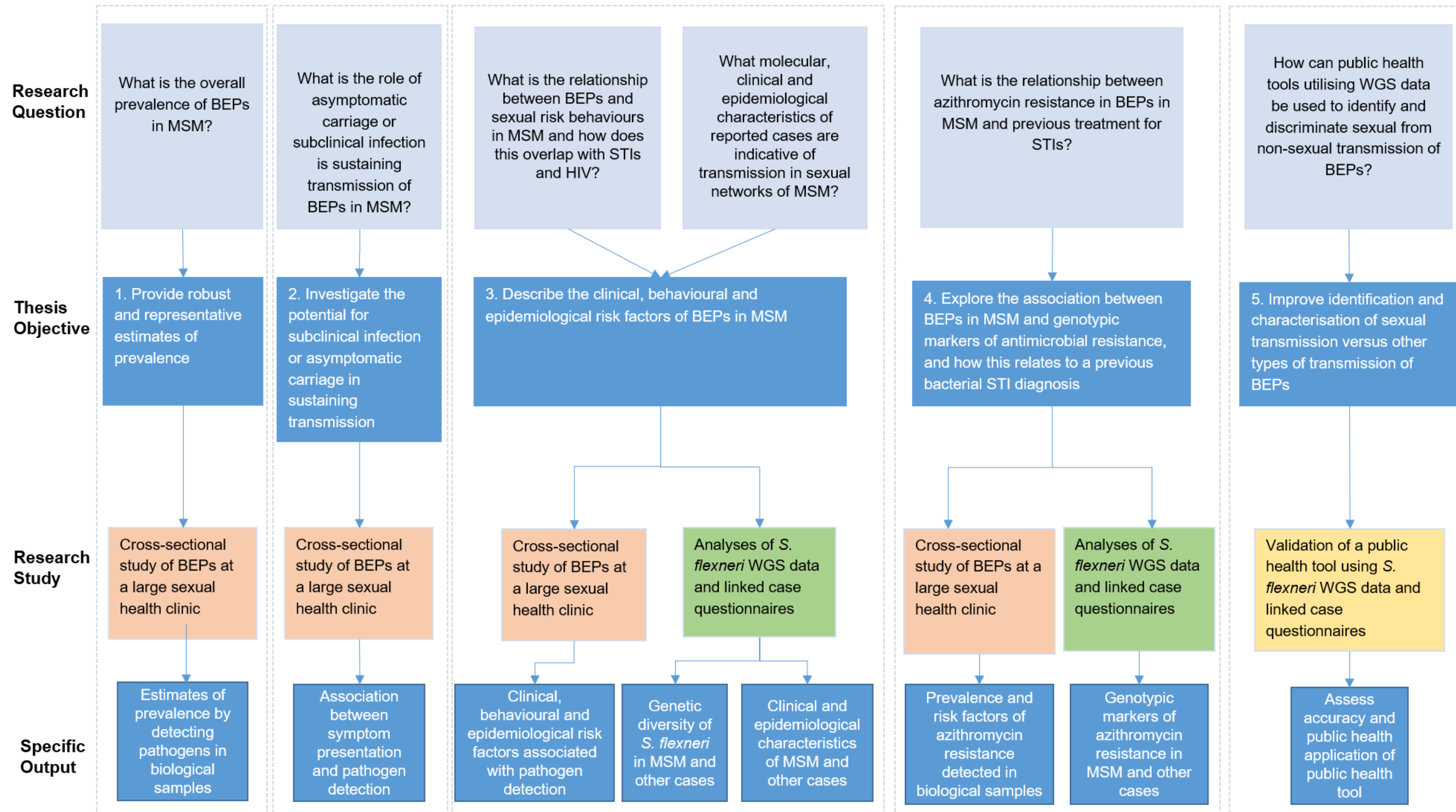


Figure 1.1: Schematic representation of the framework for this thesis

1.5 My role in this research

Chapter 3 describes a cross-sectional study of BEPs among MSM attending a large SHC in central London. I led the design, development and implementation of this study, including the NHS ethics, Heath Research Authority (HRA) and local research and development (R&D) approval processes. I arranged shipping of specimens to the Gastrointestinal Bacteria Reference Unit (GBRU) at Public Health England (PHE), where I carried out all laboratory procedures. I undertook all data analyses. The contributions of others, according to the study protocol and described further in Chapter 3, are as follows:

- A research nurse, Lorraine Omari-Asor, from the National Institute for Health Research Clinical Research Network (NIHR CRN) managed the study opt-out log, and generated the list of participants who opted-out and the list of participants who were eligible for the study,
- Two research nurses, Alexandra Schoolmeesters and Serah Duro, from Chelsea and Westminster NHS R&D Department extracted patient records from the clinic patient database.
- Clinical staff at the SHC, Dr Gary Whitlock, Dr Jey Zdravkov and Jenny Olsson, extracted clinical records from the clinic patient database.
- A surveillance scientist within the HIV & STI Department at PHE (initially Paula Blomquist, subsequently Megan Bardsley) extracted GUMCAD STI Surveillance System (formerly known as the Genitourinary Medicine Clinic Activity Dataset and herein referred to as GUMCAD) records and linked these to the study participants. In addition, the PHE scientist was also responsible for anonymising the data prior to my analysis.
- GBRU Standard Operating Procedures (SOPs) were used to perform all laboratory procedures including DNA extraction and PCR testing. I received training and guidance from Panida Silalang.

Chapters 4 and 5 present analyses of *S. flexneri* WGS data and epidemiological data collected as part of national surveillance at PHE. I spent time with the Bioinformatics Team at the GBRU, where I received training and support with the analysis and interpretation of WGS data. The collection and processing of specimens, WGS and preliminary analysis of raw sequencing

data, were carried out by PHE according to standard protocols and are described in detail in Chapters 4 and 5. I performed all data cleaning and analyses, including phylogenetic analyses. The contributions of others, described further in Chapters 4 and 5, are as follows:

- I carried out phylogenetic analyses using standardised protocols developed by the Bioinformatics Team at PHE (i.e. SnapperDB). Training and support, including help with interpretation, were provided by Anaïs Painset and Dr Tim Dallman (Chapters 4 and 5).
- Epidemiological data were collected as part of a pilot of a new standardised exposure questionnaire. Data were collected by PHE as part of public health follow-up of cases. Data entry was performed by two members of the HIV & STI Department at PHE, Tracey Cairns and Krishna Gupta (Chapters 4 and 5).
- A surveillance scientist within the HIV & STI Department at PHE, Peter Kirwan, extracted HIV surveillance system records and linked these to the combined WGS and epidemiological dataset according to a standardised protocol. The PHE scientist irreversibly anonymised all data before I performed the analyses (Chapter 4).
- I used a standardised computer programming script developed by the Gastrointestinal Infections Department at PHE to classify transmission clusters. Training and support were provided by Amy Mikhail (Chapter 5).

Chapter 2: Background

This chapter sets out the context and rationale for my PhD research. Firstly, I introduce the background to enteric pathogens in England, including the surveillance systems in place for identifying cases and the public health strategies used for prevention and control. I then describe what is already known about the epidemiology of enteric pathogens in MSM in England and elsewhere, including the characteristics associated with recent outbreaks, and the key issues to be addressed for improved prevention and control.

2.1 Introduction to enteric pathogens

Enteric pathogens can cause disease of the intestinal tract. They commonly cause gastroenteritis (inflammation of the stomach and intestines), which is characterised by the sudden onset of diarrhoea and/or vomiting, with additional symptoms including abdominal pain and fever. In severe cases, significant morbidity and mortality may be associated with dysentery (bloody diarrhoea), severe dehydration, bacteraemia, weight loss or haemolytic uremic syndrome (HUS), characterised by anaemia and acute kidney failure.[29-31] Some enteric pathogens cause systemic infection but few symptoms of gastrointestinal illness.[32] Furthermore, asymptomatic carriage or subclinical infections can occur.[32-34]

Transmission occurs via the faecal-oral route, either through direct physical contact with an infected person, through exposure to environmental sources (e.g. contaminated surfaces, toilets or objects), the consumption of contaminated food or water, or through contact with animal reservoirs or their environment.[32,35,36] As well as being primary sources of infection, food and water can also become contaminated via faeces from an infected person or animal. In the UK, many cases are linked with foreign travel to low-income regions with poor food and water hygiene, such as South Asia or sub-Saharan Africa, where enteric pathogens are endemic.[1,2,35,36] Sexual transmission, particularly in men who have sex with men (MSM), can occur through direct oral-anal contact (i.e. rimming), or through oral sex after sex, or via fingers or fomites.

2.2 Aetiology and clinical features

There are a range of bacterial, protozoal and viral agents that can cause an enteric infection (Table 2.1). The likelihood of acquiring one of these depends on several factors related to the pathogen, the host and the environment.[37] For example, the infectious dose, a measure of the number of organisms required to establish an infection, varies by pathogen. Those with a very low infectious dose, such as *Shigella* spp., are highly infectious and frequently spread through person-to-person transmission.[31,38,39] With regards to host-related factors, gastrointestinal illness is more common and likely to be of greater severity in specific groups of people.[29,31] These include people at extremes of age (i.e. those over 60 years old and young children), people with pre-existing medical conditions that may alter the immune response, either because of the disease itself (e.g. HIV) or the use of medication (e.g. recipients of transplants taking immunosuppressant drugs), people who are malnourished, and people taking medications to reduce stomach acidity (e.g. proton-pump inhibitors).[29,31,40] Recent antimicrobial use may also indicate altered intestinal flora and decreased ability to resolve infection.[29,31] Environmental conditions such as private drinking water supplies or living in a rural area with frequent exposure to animals (or their faeces) may increase the likelihood of infection.[41,42]

The symptoms and clinical signs of infection vary depending on the infecting pathogen. In most cases, clinical symptoms tend to be acute (typically 5-7 days and less than 14 days) and self-limiting.[29,31] Persistent diarrhoea (more than 14 days) is more likely with protozoan pathogens such as *Giardia lamblia*, *Entamoeba histolytica* and *Cryptosporidium* spp.[26,31] Bloody diarrhoea in febrile patients is usually associated with invasive pathogens such as *Shigella* spp., *Campylobacter* spp., *Salmonella* spp., or *E. histolytica*.[30]

Shigella spp. are the most common bacterial cause of severe bloody diarrhoea.[2,43,44] The *Shigella* genus comprises four different species: *S. boydii*, *S. dysenteriae, S. flexneri* and *S. sonnei*. The first three species are further sub-divided into serotypes based on the structure of the O-antigen, a major component of the surface lipopolysaccharide (LPS) of Gram-negative bacteria.[44] *S. sonnei* cannot be serotyped as it has only one LPS O-antigen.[45] *S. flenxeri* and *S. sonnei* are responsible for the greatest burden of shigellosis globally and in England.[44,46]

S. sonnei causes a relatively mild form of shigellosis characterised by watery or bloody diarrhoea (Table 2.1). *S. flexneri* infection is characterised by bloody diarrhoea, fever and abdominal pain which can be more prolonged and of greater severity than shigellosis caused by *S. sonnei*. S. *flexneri* serotype 2a causes a substantial burden of shigellosis that has previously been associated with a specific enterotoxin (shET1) that is not commonly found in other *S. flexneri* serotypes.[44,47] *S. boydii* causes shigellosis of varying severity, but often causes bloody diarrhoea. The most severe form of shigellosis is caused by infection with *S. dysenteriae* type 1 due to the production of a specific toxin (Shiga toxin) that can result in HUS.[30] Severe bloody diarrhoea occurs in most cases and hospitalisation rates are higher compared to infections with other *Shigella* spp.[48]

Shiga toxin-producing *Escherichia coli* (also referred to as verocytotoxin-producing *E. coli* (VTEC) but herein referred to as STEC), an enterohaemorrhagic *E. coli* (EHEC), is also capable of producing a toxin that is closely related to the Shiga toxin produced by *S. dysenteriae* type 1.[30] Although relatively rare (916 cases in England in 2017), STEC is the most clinically important diarrhoeagenic *E. coli* strain due to its ability to cause life-threatening disease and the low infectious dose (10-100 organisms).[39]

There are some pathogens that cause systemic illness but few symptoms of gastrointestinal illness, such as hepatitis A virus (HAV) and *Salmonella* Typhi or Paratyphi A (typhoid or paratyphoid fever).[30,32]

Bloody diarrhoea, fever, persistent diarrhoea, severe dehydration or weight loss are indicative of severe infection and may require hospitalisation.[29,31] The risk of developing further complications and the need for treatment increase if symptoms are prolonged. Although rare, severe gastrointestinal symptoms and high fever may indicate the presence of bacteria in the blood.[29,31] Once in the bloodstream, the bacteria can spread throughout the body causing significant morbidity and mortality.

Table 2.1: Selected enteric pathogens of public health importance in England

Causative agent	Infectious dose	Usual incubation period	Usual duration of symptoms	Clinical features	Complications	Usual modes of transmission
Bacterial						
Campylobacter spp.	Usually 10,000, but may be as low as 500[49,50]	2-5 days	2-7 days	Abdominal pain, watery or bloody diarrhoea, fever. Asymptomatic in 25-50%	Guillain-Barré syndrome (<2%), Reactive arthritis (<10%)[51]	Foodborne, predominantly undercooked poultry or unpasteurised milk
Shiga toxin-producing *Escherichia coli* (STEC)	10-100	2-4 days	1-10 days	Ranges from asymptomatic to severe bloody diarrhoea and abdominal pain. Often without fever	HUS (5-14%)[52]	Foodborne, particularly undercooked meat. Person-to-person. Animal contact
Shigella spp.	10-100[38]	1-3 days (up to 7 days for *S. dysenteriae*)	4-5 days for *S. sonnei* Up to 7 days for other *Shigella* spp.	*S. sonnei:* Mild illness in most cases. Abdominal pain, watery (sometimes bloody) diarrhoea. *S. boydii, S. dysenteriae S. flexneri:* Generally, more severe bloody diarrhoea, fever, abdominal pain. *S. dysenteriae type 1*: Serious disease and prolonged illness	Reactive arthritis Reiter's syndrome Toxic megacolon HUS (*S. dysenteriae* serotype 1 only)	Person-to-person

Causative agent	Infectious dose	Usual incubation period	Usual duration of symptoms	Clinical features	Complications	Usual modes of transmission
Salmonella spp. (non-typhoidal)	Usually 1,000-100,000 but may also be as low as a few organisms	1-2 days	4-7 days	Watery or bloody diarrhoea, abdominal pain, fever, myalgia, headache	Septicaemia with abscess formation	Foodborne
Protozoal						
Giardia lamblia	Usually 100, but may be as few as 10	5-16 days	Variable Average 2-3 weeks	Diarrhoea, malaise, flatulence, bloating, weight loss. Often asymptomatic.		Waterborne Person-to-person
Entamoeba histolytica	1-100	2-4 weeks	Variable	Asymptomatic in 90%. Bloody diarrhoea (amoebic dysentery), abdominal pain.	Extra intestinal disease in <1% (e.g. liver abscess)[53]	Waterborne Person-to-person
Cryptosporidium spp.	10-100	4-7 days	2 days - 4 weeks	Watery or mucoid diarrhoea, bloating, abdominal pain. Often asymptomatic	Prolonged and severe illness in immunosuppressed people	Waterborne Person-to-person
Viral						
Norovirus	1-100	1-2 days	1-5 days	Nausea, vomiting, watery diarrhoea. Often asymptomatic		Person-to-person
Rotavirus	10-100	2-4 days	1-3 days	Watery diarrhoea, fever, vomiting		Person-to-person
Hepatitis A virus	10-100	Average 28 days	1-2 weeks	Fever, nausea, malaise, jaundice. Can be asymptomatic	Acute liver failure Relapsing hepatitis (up to 1 year)	Person-to-person

Adapted from: Hawker *et al.* (2012)[48] and Public Health England (2020)[46]. Additional sources: DuPont *et al.* (1989),[38] Black *et al.* (1988),[49] Janssen *et al.* (2008),[50] Esan *et al.* (2017),[51] Tarr *et al.* (2005)[52] and Haque *et al*. (2003).[53] Person-to-person transmission refers to direct physical contact with an infected person or indirect contact via contaminated surfaces or food items. *Giardia lamblia* is also known as *Giardia intestinalis* or *Giardia duodenalis*.

2.3 Diagnosis and clinical management

In the following situations, patients presenting to healthcare services with symptoms of gastroenteritis may be asked to provide faecal samples, which undergo microbiological testing to identify the causative agent(s) and in some cases, assess antimicrobial susceptibility:[54]

- Blood, mucous or pus in the stool
- Persistent diarrhoea/malabsorption
- Systemic illness with a history of diarrhoea and/or vomiting
- Recent hospitalisation
- Recent use of antimicrobials
- History of foreign travel
- Immunosuppressed

For bacterial agents, the current UK Standards for Microbiology Investigations recommend that primary diagnostic laboratories perform routine bacterial culture on all diagnostic samples for *Campylobacter* spp., *Salmonella* spp., *Shigella* spp., and STEC serogroup O157, the most common serogroup of STEC in England. Diagnostic laboratories cannot routinely isolate non-O157 serogroups.[39,54] Molecular diagnostic methods based on Polymerase Chain Reaction (PCR) assays are becoming more widely available in primary diagnostic laboratories as a rapid diagnostic test prior to performing bacterial culture. Further bacterial pathogens may be added to the standard range of diagnostic tests depending on the specific circumstances of the case, for example travel-associated diarrhoea, immunosuppression, type and duration of symptoms or hospitalised individuals.[39,54] For intestinal parasitic infections, the guidelines recommend that all cases presenting with acute diarrhoea are tested for *Cryptosporidium* spp.[55] Faecal samples can be tested for additional infections (known as the 'ova, cysts and parasites test') depending on the circumstances of the case, which may require up to three different stool samples taken two to three days apart.[39,54] The diagnosis of parasitic infections is mainly by microscopy, which is both low in sensitivity and specificity. However, larger laboratories may use antigen detection by performing enzyme immunoassays (EIA) followed by further validation.[55-57] PCR-based assays for detecting parasitic infections are not yet widely

available in primary diagnostic laboratories.[56] For viruses, faecal samples are not routinely tested except in certain circumstances, for example in food-handlers, young children or the immunosuppressed.[54] Where indicated, viruses are usually detected by EIA or PCR.[39] Recent hepatitis A infection is diagnosed by the detection of HAV-specific antibodies in the blood (IgM).[58] HAV RNA detection by Reverse Transcriptase-PCR (blood or faeces) is more sensitive but is not widely available. Primary diagnostic laboratories can refer samples to the specialist and national reference laboratories, which provide a comprehensive range of services including pathogen detection, species identification and molecular typing.[59]

Acute and self-limiting enteric infections are generally managed conservatively with oral rehydration therapy if necessary.[29,60] Antimicrobials are not usually recommended, particularly where the aetiology is unknown. In mild cases, the risk of antimicrobial side effects outweighs the benefits and improper use contributes to antimicrobial resistance (AMR).[29,31,61,62] For some pathogens, antimicrobials may result in prolonged faecal shedding (*Salmonella* spp.)[63] or increase the likelihood of HUS (STEC).[29,31,64-66] However, antimicrobial treatment may be appropriate in very severe cases or if the individual is at higher risk of complications due to immunosuppression or comorbidities.[26,29] Microbiological confirmation and antimicrobial susceptibility testing are preferable to guide the choice of antimicrobial, but empirical treatment may be necessary and should be based on microbiological advice and local susceptibility data. Typical treatment options usually include fluoroquinolones (e.g. ciprofloxacin or levofloxacin) or macrolides (e.g. azithromycin or erythromycin). Antimicrobials are not used for the treatment of STEC.[67] Specific treatment is available for diarrhoeal illness caused by protozoan pathogens including *G. lamblia* (metronidazole) and *E. histolytica*. Metronidazole is prescribed for amoebic dysentery and amoebic liver abscesses, followed by diloxanide furoate. The latter is also provided as a single drug regime for asymptomatic patients with *E. histolytica* cysts present in the faeces to prevent the infection from progressing to invasive disease.[68]

2.4 Detection and surveillance of enteric pathogens in England

Infectious disease surveillance consists of the systematic and ongoing collection, analysis and dissemination of data for public health purposes.[69] These data are used to monitor epidemiological trends, to detect outbreaks, to inform the planning and delivery of interventions

or new policies, and to provide evidence for evaluation activities. When new threats or emerging problems are detected, rapid and appropriate action can be taken to protect the health of the public. In England, surveillance for enteric pathogens draws on three main sources of data, as outlined in Table 2.2.

Table 2.2: Data sources for the national surveillance of enteric pathogens in England

Source	Description
Clinical case reports	• Statutory notifications from registered medical practitioners • Additional cases identified through routine follow-up of cases (sporadic cases or during outbreak investigations) • Voluntary reports from members of the public
Laboratory reports	• Statutory reports of notifiable organisms • Voluntary reports from diagnostic laboratories • Data from specialist and national reference laboratories
Reports of outbreaks	• Reports of foodborne and non-foodborne outbreaks, defined as either two or more cases of the same infection that are linked to the same source, or where the observed number of cases exceeds the expected number of cases and the same source is suspected. Non-foodborne outbreaks include those linked to recreational water exposure, environmental exposure at outdoor events, contact with animals or their faeces or STEC outbreaks spread through person-to-person transmission.

Source: Tam *et al.* (2012),[32] Public Health England (2013)[70] and Public Health England: Gastrointestinal infections: guidance, data and analysis.[71] Contains public sector information licensed under the Open Government Licence v3.0.

In England, 'The Health Protection (Notification) Regulations 2010' set out the legal duties of registered medical practitioners (RMPs) with regards to reporting suspected cases of certain infectious diseases (known as notifiable diseases), and the legal duties of diagnostic laboratories with regards to reporting certain infectious organisms (known as causative agents).[25] The primary purpose of these regulations is to enable timely investigation, risk assessment and public health action to control the further spread of an infection that may pose a significant risk to human health. Secondary to this, statutory notifications provide a timely source of data for public health surveillance.[72] All statutory notifications are collated by Public Health England (PHE), and analyses to describe local and national trends are performed and published on a weekly basis.

Notifiable diseases include infectious bloody diarrhoea, HUS, food poisoning and acute infectious hepatitis. RMPs have a statutory duty to notify the 'proper officer' of the relevant Local Authority (LA) (i.e. local government area) for the area in which they attend a suspected case. This 'proper officer' may be a senior environmental health officer (EHO) within the LA, or a consultant in health protection or communicable disease control within the local Health Protection Team (HPT) at PHE. HPTs are local PHE infection control teams (21 in total in England) providing public health advice and operational support for infectious disease outbreaks. If the 'proper officer' is based within the LA, they are required to forward the notification to the local HPT.[72] All suspected cases of a notifiable disease must be reported to the HPT within three days or as soon as possible (and always within 24 hours) for urgent cases. Determining whether a case requires urgent notification depends on the nature of the disease, the route of spread, ease of transmission, and the specific circumstances of the case. The key consideration is whether prompt public health action and intervention are likely required to reduce the impact on human health and to prevent further spread. With regards to the notifiable diseases listed at the start of this paragraph, all are likely to be urgent, except for food poisoning, which is considered urgent only when there is a suspected cluster or outbreak, or if an individual poses an increased risk of spreading the infection to others (e.g. food handler).[36,72] In practice, gastrointestinal illness may be caused by infectious or non-infectious agents, however, RMPs should not wait for laboratory confirmation before notifying a suspected case of a notifiable disease.[72] The algorithm for the notification of infectious diseases (NOIDS) by RMPs is presented in Figure 2.1.

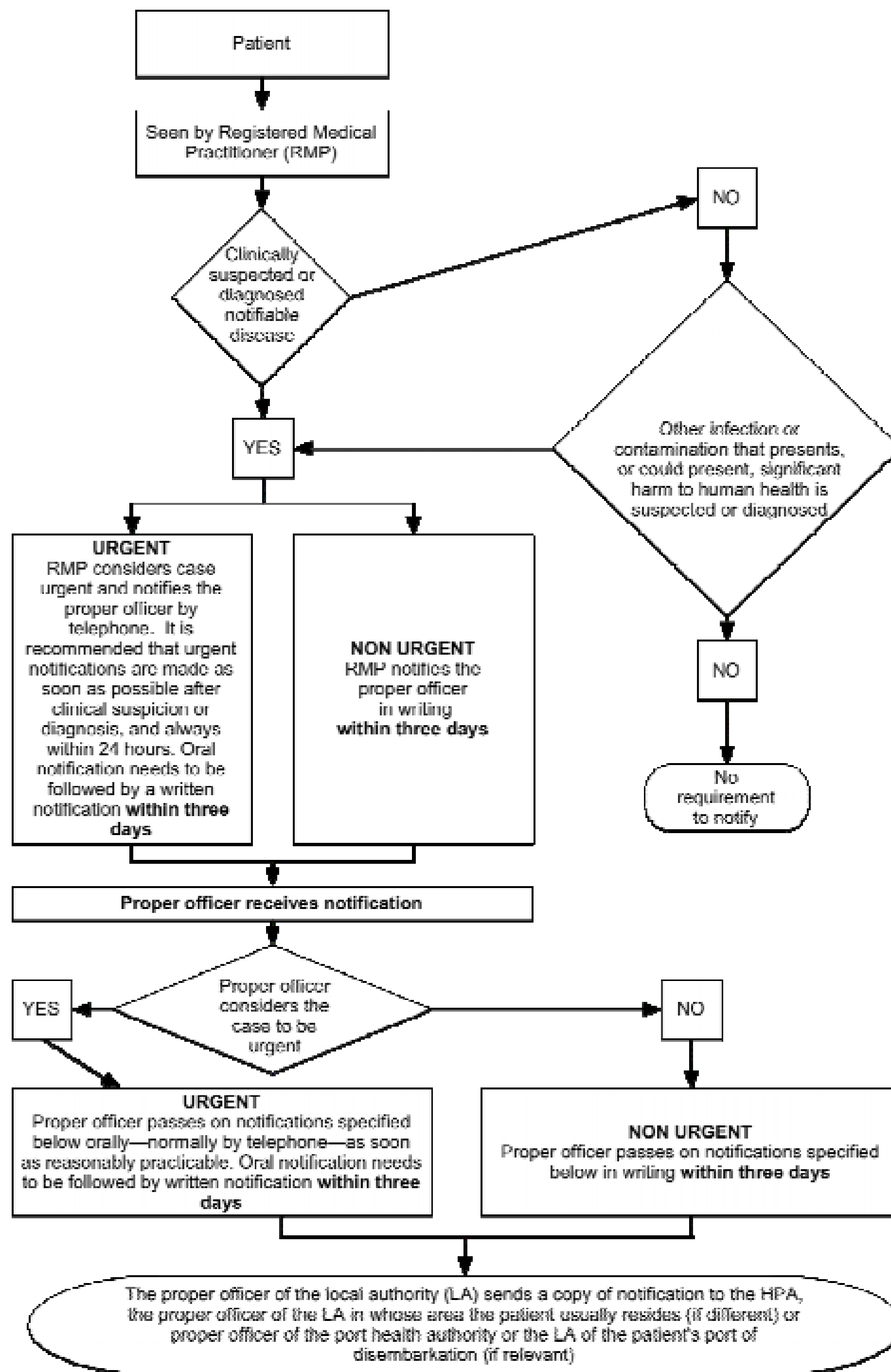


Figure 2.1: Notification of Infectious Diseases (NOIDS) by registered medical practitioners

Source: Department of Health (2010)[72] Contains public sector information licensed under the Open Government Licence v3.0

All laboratories performing a primary diagnostic role are statutorily required to notify the HPT upon identifying a causative agent. This includes nearly all enteric pathogens and these reports can verify (or refute) a clinical diagnosis and influence the public health action taken. Causative agents must be reported to PHE within seven days of identification or within 24 hours for urgent cases.[72] Examples of enteric pathogens that are likely to require urgent notification include STEC, *Shigella* spp. (except *S. sonnei*), hepatitis A virus, *S.* Typhi, *S.* Paratyphi A, *Listeria monocytogenes* and *Vibrio cholerae*. Other enteric pathogens may be considered urgent if there is a suspected outbreak or cluster or if the case is at increased risk of spreading their infection to others (e.g. food handler). For some cases, the RMP may have notified the HPT based on clinical suspicion of a notifiable disease, however, the diagnostic laboratory is still required to notify the causative agent.

In addition to statutory reports from laboratories, there are two further sources of laboratory reporting data that are collected for public health surveillance purposes. The first source of data is the national laboratory reporting database, known as the Second Generation Surveillance System (SGSS), which captures information on infectious disease agents isolated at laboratories across England and Wales.[73] SGSS is a voluntary system that includes data submitted by diagnostic, regional public health and PHE national reference laboratories. The database includes a broader range of pathogens in comparison to the list of notifiable causative agents. Reports of laboratory confirmed cases are published by PHE on a weekly or monthly basis. *Campylobacter* spp. are the most commonly isolated enteric pathogen in England with 96.6 cases per 100,000 population reported in 2017 (Table 2.3).

The second source of laboratory data includes information generated by PHE national reference laboratories for samples that are referred for species identification and molecular typing. These data are utilised for public health surveillance and outbreak detection, as well as for research. For example, whole genome sequencing (WGS) data from all cultured isolates of *Shigella* spp., *E. coli*, *Salmonella* spp. and *L. monocytogenes* referred to the Gastrointestinal Bacterial Reference Unit (GBRU) are used to monitor epidemiological trends, to support outbreak investigations and more recently, for real-time identification of clusters of cases. Data generated by the GBRU are stored within a database known as the Gastro Data Warehouse

(GDW). These data are explored and analysed in depth in Chapters 4 and 5 of this thesis within the context of identifying and characterising sexual transmission of *S. flexneri* in MSM to inform the public health response.

Table 2.3: Laboratory confirmed cases of selected enteric pathogens in England, 2017

Causative agent	Laboratory confirmed cases	Rate per 100,000 population
Campylobacter spp.	53,395	96.6
Salmonella spp. (non-typhoidal)	8,664	15.7
Norovirus	5,167	9.4
G. lamblia	4,702	8.5
Cryptosporidium spp.	4,032	7.3
Rotavirus	3,451	6.2
Shigella spp.	1,928	3.5
Hepatitis A virus	899	1.6
STEC O157	532	0.96
STEC non-O157	384	0.69
E. histolytica	43	0.08

Data from: Public Health England, Second Generation Surveillance System (SGSS)

Surveillance data have limitations, however, because they only represent a proportion of all people who acquire an infection. Some people will not develop obvious clinical symptoms, and of those that do, not all will present to healthcare.[74] Furthermore, stool samples for microbiological investigation are only requested in a subset of patients, and where requested, some people do not provide samples due to embarrassment, concerns around hygiene or lack of instructions on how to collect the sample.[75] Where stool samples are provided for diagnostic testing, not all have a pathogen detected, and where a pathogen is detected, the result is not always reported to national surveillance.[74,76] Moreover, only a sub-set of reported isolates are referred to the appropriate national reference laboratory for species identification and typing Therefore, routine surveillance data underestimate the true burden of infection in the population. This is often presented in the form of a surveillance pyramid, which describes the different levels that contribute towards the overall burden of disease in the community (Figure 2.2). It is commonly perceived that the tip of the pyramid is likely to represent those patients with the most severe symptoms.[74]

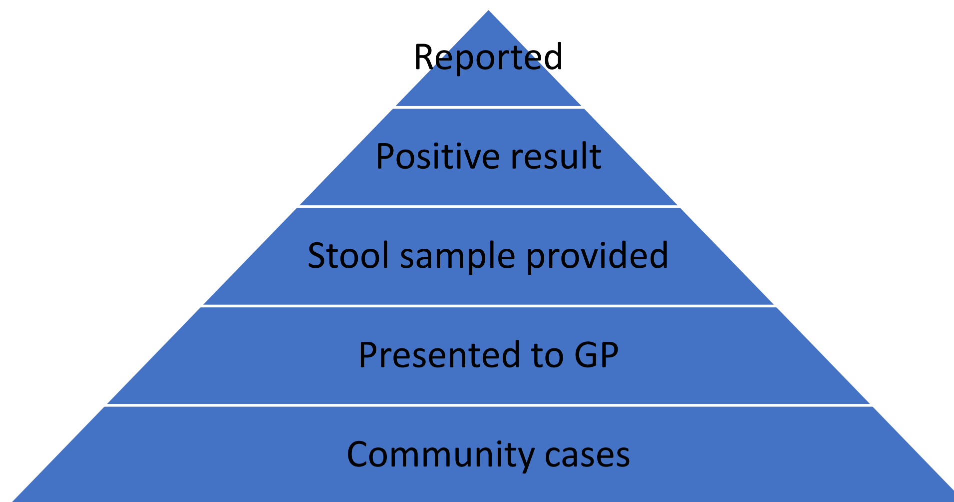


Figure 2.2: The surveillance pyramid

Adapted from O'Brien *et al.* (2010)[74]; Open Access article distributed under the terms of the Creative Commons Attribution License (CC-BY 2.0)

2.5 Incidence of enteric infections in the UK

Two UK studies, known as the first and second studies of 'Infectious Intestinal Disease (IID) in the Community', have estimated the incidence and aetiology of enteric infections among people of all ages at a population level.[74,76-78] The key components in both studies were: (i) a population-based cohort study to measure incidence and aetiology in the community, (ii) a prospective study of people who consulted their GP with gastrointestinal symptoms, and (iii) a national surveillance study to measure the ratio of reported cases to cases in the community. Cases were defined as people who had loose stools or clinically significant vomiting (i.e. more than once in a 24-hour period, incapacitating or accompanied by other symptoms such as abdominal cramps or fever) in the last 14 days, preceded by a three-week period of no symptoms.[74] People with a known non-infectious cause were excluded (e.g. Crohn's disease). Participants were provided with stool sample kits and requested to submit samples for microbiological testing for a comprehensive range of enteric pathogens, including bacteria, viruses and parasites. Importantly, the first IID study (IID1) also included asymptomatic controls (i.e. healthy people who did not develop symptoms) to estimate the prevalence of asymptomatic infection.

IID1 took place in England during 1993-1996 across 70 GPs selected to be representative of all practices nationally in terms of geographical location, urban/rural and socioeconomic characteristics.[77,79] Stratified random sampling based on age and sex was used to identify potential participants for the population-based cohort study. People who were enrolled were followed up for six months and asked to complete weekly diary cards about any symptoms of diarrhoea and/or vomiting (or lack of). Incident cases with symptoms were age-sex matched to an asymptomatic control systematically selected from the cohort and both were asked to complete a questionnaire and to submit a stool sample. The GP study took place in a sub-set of GPs (34 out of 70) over a 12-month period. People who presented to their GP with symptoms of diarrhoea and/or vomiting were invited to participate, and those who were successfully enrolled were age-sex matched to an asymptomatic control (identified from the practice register and subsequently invited to participate) and both were asked to provide a stool specimen and to complete a questionnaire. To estimate the degree of underreporting at a national level, incidence rates overall and for each pathogen were compared to those generated from the national laboratory surveillance system during the study period.

IID1 estimated that 20% of the population suffered an episode of diarrhoea and/or vomiting within a year, equating to 9.4 million cases per year.[78,79] Among cases in the community (i.e. cohort study cases), a pathogen was detected in 36.9% of submitted samples using conventional techniques (e.g. bacterial culture).[34] Viruses were the most commonly identified pathogens, particularly norovirus (7.0% [95% CI: 5.2% to 9.1%]), and *Campylobacter* spp. were the most commonly detected bacterial pathogen (4.2% [95% CI: 2.9% to 5.9%]). Among cases who presented to their GP, a pathogen was detected in 54.9% of submitted samples and *Campylobacter* spp. were the most commonly detected (12.2% [95% CI: 11.0% to 13.5%]). The detection of a pathogen in samples from asymptomatic controls was low in comparison to cases in both the community and GP study components at 16.8% and 19.0% of specimens, respectively. *Campylobacter* spp. were detected in 0.7% (95% CI: 0.2% to 1.8%) of community specimens and 0.7% (95% CI: 0.4% to 1.1%) of GP specimens. No asymptomatic controls had *Shigella* spp. detected.[34] For every diagnosis of a gastrointestinal infection recorded in the national laboratory data, there were approximately 23 cases presenting to their GP and 136 cases within the community.[78] This varied by pathogen, for

example, the ratio of cases in the community to those reported to national surveillance was 1500 to 1 for norovirus, 8 to 1 for *Campylobacter* spp. and 3 to 1 for *Salmonella* spp.[78]

IID1 was the first study to estimate the incidence and aetiology of enteric infections among a large, representative sample in England. The study raised awareness of the incidence and public health impact of gastrointestinal infections, particularly with regards to the large number of symptomatic cases within the community that do not appear in national surveillance data, and the underlying prevalence of asymptomatic carriage.

The second study (IID2) took place during 2008 to 2009 at 88 GPs across the whole of the UK.[74] The study design was similar to IID1, except that stool samples were only collected from people who developed symptoms. In addition, microbiological testing was performed using both conventional methods and PCR (the latter was not used during IID1). Potential participants for the cohort study were selected at random from the age-sex registers of the GPs. Enrolled individuals were followed up on a weekly basis for 12 months to find out if they had any symptoms of diarrhoea and/or vomiting (n=6836). Those who developed symptoms were asked to provide a stool sample and to complete a questionnaire. The GP study took place in half the GPs (37 out of 88). People presenting to their GP with symptoms of diarrhoea and/or vomiting were invited to participate in the study and those who were recruited were asked to provide a stool sample and to complete a questionnaire (n=991).[74,76] To estimate the degree of underreporting, incidence rates overall and for each pathogen were compared to those generated from the national laboratory surveillance systems. Surveillance data included any diagnosed enteric pathogen that was tested for in IID2.[32]

IID2 estimated that 25% of the population suffered an episode of IID within a year, equating to 17 million cases per year.[76] A pathogen was detected in 40% of specimens submitted by community cases (i.e. the cohort study) and viral pathogens, such as norovirus (16.5% [95% CI: 14.0% to 19.3%]), were the most commonly detected.[80] *Campylobacter* spp. were the most commonly detected bacterial pathogen (4.6% [95% CI: 3.2% to 6.3%]) followed by EAEC (1.9% [95% CI: 1.1% to 3.1%]). Other bacterial pathogens were detected in less than 1% of cases and no *Shigella* spp. were detected. Among cases who presented to their GP, a pathogen was detected in 51% of submitted specimens. *Campylobacter* spp. were detected

in 13.0% (95% CI: 10.9% to 15.5%) of specimens and norovirus was detected in 12.4% (95% CI: 10.2% to 14.7%). For every diagnosed case of an enteric pathogen recorded in national surveillance data, there were approximately 10 cases presenting to their GP and 147 cases within the community.[76]

Compared to the mid-1990s, the rate of diarrhoea and/or vomiting in the community had increased by an estimated 43%, while the rate among people consulting their GP was 50% lower.[32,76] Reasons for the decline in GP consultations may be related to increased self-management, increased availability of alternative sources of information, such as NHS Direct, and/or changes in the severity of infection over time. However, only 2% of individuals with diarrhoea and/or vomiting used NHS Direct in the IID2 study, which would not account for the large drop in healthcare usage.[32,76]

The interpretation of the results from the IID studies requires an appreciation of the type of laboratory methods used. Retesting of archived stool samples from IID1 using PCR increased the detection of eight selected pathogens in cases and controls from 53% to 75% and 19% to 42%, respectively.[81] Of note was the high proportion of specimens in both studies that did not have a pathogen detected, which might reflect the range of pathogens included in the testing panel.[80] Additionally, the absence of a pathogen in stool specimens may reflect non-infectious causes of diarrhoea, including temporary changes in bowel movements.[76,80]

With regards to the generalisability of the IID studies, only 35% of those invited participated in the IID1 cohort study, but the characteristics of the study population were broadly comparable to those of the general population in England.[77] However, enrolment was slightly lower among men, those aged 15 to 24 years, those from a lower social class (based on earnings and occupation), and among those who were not married.[77] Participation in the IID2 cohort study was only 9% of those invited and the characteristics of the study population were not representative of the general population in terms of age and sex; teenagers and young adult males were underrepresented. The low level of participation could limit the generalisability of the findings if those who participated differed in their risk of IID or tendency to report symptoms compared to those in the general population. To adjust incidence estimates for under-ascertainment (i.e. the proportion of eligible cases that were recruited), GP databases were

searched to identify all consultations that were eligible for the study. 64% of eligible cases presenting to their GP were recruited in the GP study of IID1 compared to 17% of cases in IID2.[76]

2.6 Prevention and control

PHE HPTs, in collaboration with EHOs at the relevant LAs, are responsible for the public health follow-up and management of enteric infections.[36,46] Their role includes local level surveillance, responding to alerts or reports of any notifiable diseases or causative agents, investigation and management of outbreaks and the implementation of control measures.

Public health follow-up involves contacting the case and conducting a risk assessment to assess the risk of ongoing transmission, to investigate possible sources of exposure, to identify and manage any associated household or close contacts and to determine whether there are any connected cases or links to a known outbreak. To do this, the EHO or an individual from the HPT will interview the case using a structured questionnaire, and in some cases, contact the attending clinician to obtain further details. Some pathogens are not routinely followed-up unless they are known to be part of a cluster or outbreak (e.g. *Campylobacter* spp.). However, there are some enteric pathogens where routine public health follow-up is required for all cases due to the potential severity of infection (e.g. STEC and *Shigella* spp. [except for *S. sonnei*]). Standardised national questionnaires are used in these situations to collect epidemiological data including demographics, potential sources of infection, close contacts and details on the clinical condition.[27,46,82] The HPT also plays an important role in ensuring that microbiological test results are recorded and reported to national and regional surveillance teams. Where the infectious agent is not known, the HPT will liaise with EHOs and microbiology laboratories to ensure that this is completed to enable a prompt and appropriate public health response.[27,36,82] If an outbreak is suspected, often defined as two or more cases with the same infection that are linked in time and place, or when the number of cases is higher than expected,[70] an outbreak control team might be established and usually includes representation from the HPT, LA, NHS or public health microbiology laboratory and the PHE regional Field Epidemiology Service team.[83]

General measures taken to prevent onward transmission of enteric pathogens include advice on toileting, hand and food hygiene (Box 2.1) and exclusion from work, school or other institutional settings, until at least 48 hours after the case is symptom free.[46] For hepatitis A, the exclusion period is seven days after the onset of jaundice.[84] This advice should be provided by the attending clinician and is reinforced by the HPT (or EHO) when they contact the case. Guidelines for the public health management of gastrointestinal infections have only recently been updated to reflect the risk of transmission through sexual contact, in view of recent outbreaks predominantly affecting MSM.[46] In addition, the British Association for Sexual Health and HIV (BASHH) recently updated their national guidelines on the sexual health care of MSM who have symptoms suggestive of an enteric pathogen.[26] The guidelines recommend that all people with a confirmed or suspected enteric pathogen should be provided with advice on the prevention of non-sexual and sexual transmission of infection. Specific measures for preventing the sexual transmission of infection include advice on hygiene practices before and after sex (e.g. washing hand, genitals and perianal skin), the use of barrier protection for specific sexual risk practices (e.g. rimming), refraining from sharing sex toys and abstaining from sexual contact until at least seven days after the case is symptom free.[26]

To prevent further person-to-person transmission, specific attention is given to those belonging to a recognised risk group (Table 2.4). For instance, it may be necessary to inform and advise close contacts, or the workplace/school of the case. Microbiological clearance may also be required before returning to normal working. The guidelines on clearance samples vary depending on the infecting pathogen, species and risk group that the case belongs to.[36,46] For example, cases of *S. flexneri*, *S. boydii* and *S. dysenteriae* non-type 1 in a risk group require a single negative stool sample to be taken at least 48 hours after the case is symptom free, or 48 hours after the completion of antimicrobials (if appropriate), whichever is later.[27] For *S. dysenteriae* type 1, two consecutive negative stool samples (at least 24 hours apart) are required but no clearance sample is required for cases of *S. sonnei*.[27] Screening and exclusion of asymptomatic contacts of cases in a risk group is recommended for specific infections such as *S. dysenteriae* type 1 and STEC, particularly among children under five years of age.[27,82] Depending on the specific context of the case or outbreak, specific measures may be taken to protect public health including extended exclusion of cases or contacts, the provision of health

promotion messages and advice, a risk assessment of the hygiene facilities in a workplace (or other institutional setting), or vaccination (hepatitis A) or screening of close contacts.[27,36,46,82,84] LAs have the legal power to require or request action to be taken to protect human health if voluntary measures are insufficient. For example, they can require that a child is kept away from school or they can request that premises or their contents are disinfected or closed.[72]

Box 2.1: Hygiene advice to prevent the person-to-person spread of enteric pathogens

- Wash hands thoroughly with warm running water and soap and dry thoroughly:
    - After going to the toilet
    - After changing nappies
    - Before handling, preparing, serving or eating food
    - After handling soiled clothing or bed linen
    - After cleaning bedpans or toilets
    - After handling pets or non-domestic animals
    - After attending to any person who has diarrhoea or vomiting
- Avoid preparing food or handling food for other people until 48 hours after symptom free
- Avoid sharing towels or baths with someone who has diarrhoea or vomiting
- Avoid close contact, including sexual contact, with anyone who has diarrhoea or vomiting
- Use a flush toilet where possible. If a commode or bedpan is used, wear gloves and dispose of contents into the toilet. Wash the vessel with hot water and detergent, and allow to dry.
- Wash soiled clothing and bed linen separately from other clothes and at the highest temperature they will tolerate (for example 60°C or higher for linen). Dispose any excess faecal matter into the toilet before washing. Soaking in disinfectant is not necessary. Wipe down the outside of the washing machine with water and detergent after the linen is loaded.
- Clean spillages of faeces and vomit immediately with hot water and detergent. Disposable gloves should be worn and hands washed thoroughly afterwards.
- Clean toilet seats, flush handles, basin taps, surfaces and toilet door handles at least once daily with hot water and diluted bleach solution. Use separate disposable gloves and cloths to clean toilets.

Adapted from: Public Health England (2020)[46] Contains public sector information licensed under the Open Government Licence v3.0

Table 2.4: Risk groups for the transmission of enteric pathogens

Risk Group	Description	Additional comments
A	Any person who is unable to perform adequate personal hygiene due to lack of capacity or ability to comply OR has lack of access to hygiene facilitates.	Risk assessment regarding access to hygiene facilities should consider the availability of toilets/handwashing/hand drying facilities in a work/educational setting.
B	All children aged five years old or under who attend school, pre-school, nursery or other similar childcare or child minding groups.	For children aged 5 years and under who do not attend school, risk assessment for clearance purposes should explore potential for transmission within other settings e.g. household or attendance at parties
C	People whose work involves preparing or serving unwrapped ready to eat food (including drink)	Consider informal food handlers, e.g. someone who regularly helps to prepare food for charity and community events.
D	Clinical, social care or nursery staff who work with young children, the elderly, or other particularly vulnerable people, and whose activities increase the risk of transferring infection via the faecal-oral route.	Risk assessment should consider activities such as helping with feeding or handling objects that could be transferred to the mouth.

Adapted from: Public Health England (2020)[46] Contains public sector information licensed under the Open Government Licence v3.0

A significant component of the public health management of enteric infections is effective communication. The HPT is responsible for informing and advising any relevant organisations or health professionals concerning an individual case or outbreak with the purpose of supporting prevention and control efforts. This can include GPs or other NHS services, other HPTs, public health colleagues within LAs, PHE national surveillance teams and the communications departments of relevant organisations (NHS services, PHE, LA). With regards to sexual transmission or outbreaks among MSM, communication with BASHH and other charitable organisations such as Terrence Higgins Trust (THT), the UK's leading HIV and sexual health charity, may be appropriate.

2.7 Distribution and characteristics of enteric pathogens in MSM

Sexual transmission of enteric pathogens occurs through the ingestion of faecal matter during or after sexual activity. MSM are particularly at risk as they may engage in sexual behaviour that increases the likelihood of faecal-oral transmission. In this section, I present a scoping literature review which sets out the key literature relevant to the epidemiology of enteric pathogens in MSM. This type of literature review aims to i) map out the existing literature in the field, ii) clarify key concepts and themes, and iii) identify gaps in the evidence base.[85,86] I chose to conduct a scoping review because this methodology is commonly used to provide an overview of the literature across a broad topic area which has not yet been comprehensively reviewed. In addition, scoping reviews usually include evidence from a wide range of study types including review articles and case reports. This contrasts with systematic literature reviews, which typically address a specific research question by consolidating evidence from a smaller range of primary research studies that have been quality assessed.[86] In my review, I discuss the distribution and characteristics associated with outbreaks of enteric pathogens globally and the key issues for prevention and control. Epidemiological trends and control strategies specific to the UK are described in detail to provide further context for this thesis. The search terms I used to identify relevant studies in the literature are provided in Appendix 2.1. A summarised version of this section has been published: Mitchell H, Hughes G. Recent epidemiology of sexually transmissible enteric infections in men who have sex with men. *Curr Opin Infect Dis*. 2018;31(1):50-6. Permission to reuse and adapt content from this article was obtained from the rights holder, Wolters Kluwer Health, Inc.

2.7.1 Sexual transmission in context

The importance of sexual transmission in the spread of enteric pathogens was widely recognised in the 1970s. ‘Gay bowel syndrome’, now an outdated term, was used to refer to a range of anorectal and colon conditions, including traditional STIs, enteric infections and anorectal disorders found at an unusually high frequency in populations of MSM.[3,87] At the time, there were clinical case reports of enteric infections including *Shigella* spp.,[88-90] *Salmonella* Typhi,[91] *G. lamblia*,[92] *Entamoeba histolytica/dispar* complex[88] and hepatitis A

virus[88] among MSM, particularly in large urban areas of the USA, thought to have been acquired through oral-anal contact. In response to these case reports, several studies explored the prevalence and risk factors of enteric protozoa in MSM and found that these pathogens were more common in gay or bisexual men compared to heterosexual men or women (see section 2.7.6).[93-95] In MSM, oral-anal sex was often found to a be a significant risk factor for the detection of *E. histolytica/dispar* complex and/or *G. lamblia*, and one study reported a significant association with number of sexual partners and a history of gonorrhoea or syphilis (*E. histolytica/dispar* complex only).[95]

The environment and context in which the first reports of enteric pathogens in MSM were described has changed substantially. In the 1970s, there were major social changes taking place leading to greater sexual freedom and acceptance.[87,96] However, the onset of HIV/AIDS shortly after this and the subsequent population-level changes in sexual behaviour led to a reduction in the incidence of HIV and other STIs, including enteric infections.[7,96,97] These changes in sexual behaviour were likely sustained until the mid-1990s, after which reported high-risk sexual behaviours, STIs and shigellosis re-emerged among MSM in western industrialised countries, coinciding with the introduction of Highly Active Antiretroviral Therapy (HAART) for HIV.[6,7,98-102]

In recent years, the growing number of enteric pathogen outbreaks worldwide among MSM has been a public health concern, particularly due to the emergence of resistance to front-line antimicrobials.[22,23,45]

2.7.2 Aetiology of sexually transmissible enteric infections

There is evidence for sexual transmission of a wide variety of enteric pathogens through direct or indirect oral-anal contact (Table 2.5). These pathogens typically cause gastroenteritis which may be in the form of enteritis (inflammation of the intestine) or colitis (inflammation of the colon) resulting in symptoms such as diarrhoea or dysentery, vomiting and abdominal pain.[26,103] Symptoms can often overlap with those of traditional STIs, for example, proctocolitis (inflammation of the rectum and colon) can be caused by bacterial enteric pathogens (BEPs) or STIs such as lymphogranuloma venereum (LGV), causing symptoms such as rectal pain

and bleeding, mucoid discharge, the sensation of incomplete defaecation or altered bowel movements.[26,103,104] Mixed presentations and co-infections in MSM are not uncommon.[26]

Table 2.5: Aetiological agents of sexually transmissible enteric infections

Bacterial	Viral	Protozoan
Shigella spp.	Hepatitis A virus	*E. histolytica*
E. coli	Cytomegalovirus*	*G. lamblia*
Campylobacter spp.		*Cryptosporidium* spp.*
Salmonella spp.		*Microsporidium spp.**

Adapted from: de Vries *et al.* (2014)[103] *Usually in immunosuppressed people associated with HIV infection

2.7.3 Characteristics of enteric infections in MSM

2.7.3.1 Sexual behaviour and recreational drug use

Most of the data describing the specific behavioural characteristics of MSM infected with enteric pathogens have come from enhanced surveillance questionnaires or interviews conducted with symptomatic cases during outbreak investigations. These have found that the characteristics of MSM diagnosed with enteric infections are broadly similar; men report high numbers of sexual partners (e.g. up to 10 partners in the week before symptom onset was reported during an outbreak of *S. sonnei* in Berlin),[4,7,13,23] attending on-site sex venues[6,105,106] or private sex parties,[13,23] and/or the use of the internet[7] or geospatial networking applications[4,13] to meet casual partners. Networking websites and applications may have increased opportunities for sexual mixing, and to explore and experiment in different sexual behaviours without emotional risks or attachments.[107]

In some MSM outbreaks, case-control studies have been conducted to more formally establish the risk factors associated with transmission in MSM. For hepatitis A, specific behavioural risk factors have included attending on-site sex venues,[105,106] direct or indirect oral-anal contact with anonymous partners, having more than one anonymous sex partner and engaging in group sex.[108] Following an outbreak of *S. sonnei* in Sydney in 2000, a case-control study conducted among MSM found that compared to controls who had no history of diarrhoea in the previous three months, visiting a sex venue in the past two weeks was significantly associated with *S. sonnei* infection (Odds Ratio (OR) 4.8 [95% CI: 1.8 to 12.6]). Environmental

contamination of sex venues, including douching equipment, was thought to have contributed to this outbreak.[6]

For some MSM, recreational drug use including chemsex (specifically, the use of crystal methamphetamine [crystal meth], gammahydroxybutrate/gammabutyrolactone [GHB/GBL] and/or mephedrone immediately before or during sex) may enhance, lengthen and/or allow for more diverse sexual experiences by reducing inhibitions and increasing feelings of euphoria and sexual arousal.[109] This can facilitate higher risk sexual behaviours including condomless anal intercourse (CAI) with casual partners,[110-113] group sex,[109,112,114] fisting,[109,115] the use of sex toys,[23,109] and scat play.[4] Social media and networking applications have increased the profile of drug use and chemsex parties,[116] which have been linked to outbreaks of enteric pathogens in the UK including *S. flexneri* serotype 3a and STEC O117:H7.[4,13]

Data from outbreak investigations suggest that men diagnosed with an enteric infection report higher risk sexual behaviours that increase the likelihood of acquiring an infection through sexual contact. In addition, these men are often co-diagnosed with an STI or have had a recent STI diagnosis, suggesting overlapping sexual networks.[4,11-13] In the UK, the behavioural profile of MSM infected with *Shigella* spp. or STEC was similar to that reported in outbreaks of LGV and infectious syphilis[117,118]

2.7.3.2 HIV/AIDS

Enteric infections in MSM are often associated with HIV.[119-122] However, the interaction between enteric infections and HIV is complex and it is currently uncertain whether the observed association reflects biological factors such as increased susceptibility in MSM living with HIV, behavioural factors that facilitate transmission of STIs and enteric pathogens in sexual networks of MSM living with HIV, or a combination of these.[119,123]

Early clinical case reports described severe illness associated with *Salmonella* spp.[124-126] and *Shigella* spp.[127-129] among MSM living with HIV including bacteraemia and/or recurrent or relapsing infection. Most of these cases were reported prior to the introduction of HAART and were primarily observed in people who had AIDS and were immunosuppressed. The main limitation of these reports was that they were based on small numbers of cases and did not

include clinical cases in MSM who did not have AIDS. Nonetheless, it is possible that long-term shedding due to chronic or relapsing infection in HIV-immunocompromised people could contribute towards ongoing transmission in sexually active MSM.

Nelson *et al*. (1992) performed a retrospective clinical review of MSM living with HIV who presented to a London sexual health clinic (SHC) with diarrhoea between 1985 and 1991 and were diagnosed with shigellosis (n=7), campylobacteriosis (n=30) or salmonellosis (n=42). Nearly all cases of campylobacteriosis had a diagnosis of AIDS, compared to half of salmonellosis cases and only two cases of shigellosis.[130] Salmonellosis cases with AIDS were more likely to relapse or have septicaemia compared to those without AIDS. In the same study, nearly all cases of campylobacteriosis occurred in men who had a diagnosis of AIDS and in those with the lowest CD4 cell counts, but there was no association between a diagnosis of AIDS and invasive disease. On the other hand, shigellosis was diagnosed in men with higher CD4 counts, which the authors proposed was related to sexual transmission.[130] More recent clinical case reports support this theory, describing invasive shigellosis in MSM with well-controlled HIV, resulting in hospital admissions and complications such as bacteraemia or acute kidney injury associated with hypovolaemic shock.[131,132]

Laboratory-based studies have also provided insight into the increased risk of enteric infections among people diagnosed with AIDS. Laboratory records in San Francisco indicated that the annual incidence of salmonellosis was 20 times higher among adult men diagnosed with AIDS compared to adult men without AIDS.[133] Similarly, laboratory-confirmed cases of campylobacteriosis in Los Angeles showed that the annual incidence was 39 times higher among people with AIDS compared to the general population.[134] The use of laboratory records has some limitations including that people with HIV may be more likely to attend healthcare and have stool specimens collected for microbiological investigations, which could have accounted for some of the increased risk observed.[133,135] Prior to the introduction of HAART, clinical case reports suggested that campylobacteriosis was more common among MSM with advanced HIV infection,[130,135] but infections have been reported in HIV-diagnosed MSM with a range of CD4 counts, particularly within the context of an outbreak,[11,136] indicating that infections occur in MSM regardless of immune status.

In a retrospective seroprevalence study of *E. histolytica* among adults attending GPs in Sydney, HIV-diagnosed MSM with syphilis were significantly more likely to have *E. histolytic*a antibodies compared to HIV-negative MSM and a random control group of men and women who resided in the same area. The authors suggested that MSM living with HIV were more likely to develop invasive disease, however the respective roles of sexual behaviour and immune status were not explored.[137]

The role of sexual behaviour and HIV infection as risk factors for shigellosis were explored in a population-based case control study by Aragon *et al.* (2007) [122] HIV infection, sex between men, foreign travel and direct oral-anal contact were all independently associated with shigellosis among adult men reported in San Francisco between January 1998 and December 1999.[122] When restricted to MSM, foreign travel, direct oral-anal contact and HIV infection were all independently associated with shigellosis, and the proportion of cases that were attributed to these risk factors was 0.07, 0.39, 0.52, respectively. The authors suggested that increased host susceptibility because of HIV infection may increase the likelihood of shigellosis following contact with an infected person.[122] However, the study did not include any information on HIV treatment status, CD4 cell count or HIV viral load.

A national surveillance study linking laboratory confirmed reports of shigellosis in England to the national HIV database showed that between 2004 and 2015, the incidence of shigellosis in adult men living with HIV increased from 47 per 100,000 to 226 per 100,000, but remained low in adult women.[121] Among non-travel associated diagnoses of shigellosis, the proportion of men who were HIV-diagnosed was 21% compared to 2% of women. Over 90% of non-travel associated shigellosis cases living with HIV were MSM, with HIV preceding the shigellosis diagnosis in most cases (85%). In addition, where information was available, 65% had an undetectable viral load (<50 copies per ml), which might suggest that the overlap between shigellosis and HIV was more consistent with higher-risk sexual behaviours in dense networks of MSM living with HIV, rather than a function of immune status. On the other hand, UK HIV surveillance data to the end of 2015 showed that of MSM in care and receiving treatment, 94% had an undetectable viral load (≤50 copies per ml).[138] The study suggested that sexual transmission in MSM living with HIV was contributing, at least in part, to the observed increase

in shigellosis reported in England. However, the true overlap between the HIV and shigellosis epidemics may be higher as the study only included people who attended healthcare settings and whose records had sufficient data to enable linkage. 63% of shigellosis cases in this study had sufficient data for linkage.[121] Furthermore, approximately 13% of people living with HIV were undiagnosed in 2015.[139]

There is evidence to suggest that MSM living with HIV are an important group in sustaining enteric pathogen transmission in the population. However, it is important to recognise that MSM living with HIV are likely to attend health services regularly and this could artificially increase the reported occurrence of enteric pathogens in this population, particularly if clinicians are more likely to send stool samples for microbial investigations. From a biological perspective, HIV-related immunosuppression in MSM could result in increased susceptibility or long-term shedding due to chronic or relapsing infection thereby contributing towards the ongoing transmission of enteric pathogens. However, in the current era of HIV treatment, most MSM living with diagnosed-HIV are likely to be immunocompetent.[140] There are also no studies comparing clinical outcomes of enteric infections in MSM with well-controlled HIV and MSM who are HIV-negative. From a behavioural perspective, HIV-serosorting, where MSM living with HIV seek partners with the same HIV status for condomless sex, can create dense sexual networks of people living with HIV.[4,23] Although perhaps becoming less common in recent years with the widespread promotion of the U=U (undetectable=untransmittable) campaign,[141,142] HIV-serosorting may have been adopted to reduce HIV transmission, but could have facilitated the transmission of other STIs and potentially also enteric pathogens in dense sexual networks with a high prevalence of condomless sex and higher partner numbers.[20,21,143,144] In England for example, STI rates in 2013 were up to four times higher among MSM living with HIV compared to MSM who were HIV-negative or of unknown HIV status,[143] likely reflecting HIV-serosorting in sexual networks where the STI prevalence is high.

2.7.4 Recent outbreaks of public health importance

2.7.4.1 Shigella spp.

In the UK, and other high-income countries, shigellosis is typically diagnosed in travellers returning from regions with a high risk for contracting diarrhoeal disease such as sub-Saharan Africa, South Asia and Latin America,[1] but outbreaks of *S. sonnei* and *S. flexneri* have occurred among MSM in large cities around the world.[4,6,7,131,145,146] The severity of illness can be substantial, particularly with *S. flexneri*, which is more virulent and pathogenic than *S. sonnei* (see section 2.2).[48,147]

2.7.4.1.1 Recent epidemiology

In England and Wales, an increase in non-travel related *S. flexneri* serotype 3a in adult men (2009-2013) prompted a national outbreak investigation, which subsequently found that sexual transmission between men was likely driving the increase in adult male cases.[15] Until then, there had only been occasional reports of UK-acquired shigellosis associated with sexual transmission, predominantly in MSM, as well as a sporadic outbreak of *S. sonnei* among MSM in London in 2004 (n=17).[148] However, the 2009 outbreak of *S. flexneri* serotype 3a occurred across multiple geographical regions in England and Wales, and at a much larger scale than had been observed previously. Early investigations did not identify a common source or venue and the timescale of the outbreak was suggestive of person-to-person transmission.[149] Semi-structured interviews conducted with a sub-set of diagnosed men (n=34) suggested that specific sexual practices and drug-use behaviours, predominantly among MSM living with HIV, played an important role in facilitating transmission.[4] Men attended sex parties and participated in group sex to experience new sexual behaviours and some reported that chemsex drug use facilitated this. Nearly 90% of infected men had never heard of shigellosis and many visited their GP, who assumed they had food poisoning.[4] The rise in *S. flexneri* serotype 3a coincided with increases in STIs among MSM including syphilis, gonorrhoea and LGV.[20,21,150]

Since 2011, laboratory reports in England have shown subsequent waves of *S. flexneri* serotype 2a and *S. sonnei* in adult men with no recent foreign travel history, suggesting an

intensification of the shigellosis epidemic through separate introductions to the MSM population (see section 2.7.5). Similar changes in epidemiology, including shifts in the dominant serotype responsible for successive epidemics in MSM have also been observed in large cities throughout North America.[131,147,151-153] In both Montreal and Vancouver, for example, increasing rates of *S. flexneri* serotype 3a occurred among MSM from 2009 onwards, replacing the previously dominant *S. sonnei,* and coinciding with the increase in *S. flexneri* serotype 3a observed in England.[131,147]

The reasons for serotype switching may reflect levels of natural immunity in the core population.[88,131,147] Immunity to shigellosis is thought to be serotype specific so, under the right conditions, new serotypes may enter and spread within a population previously exposed to a different serotype.[154] Sufficient herd immunity could also temporarily reduce the circulation of certain *Shigella* spp. serotypes in sexual networks. Declining levels of antibodies in sexually active MSM, together with the introduction of newly susceptible individuals (e.g. as MSM start engaging in specific sexual practices), could lead to a decrease in herd immunity below a critical level resulting in renewed epidemics.[155] Cyclic epidemics of *S. sonnei* have also been observed in ultraorthodox Jewish communities and are thought to reflect natural levels of immunity and the susceptible population as people enter and leave.[155,156]

2.7.4.1.2 Public health response in England and Wales

In 2013, a targeted public health campaign was conducted to raise awareness about the sexual transmission of *Shigella* spp. among MSM and clinicians, including GPs and other healthcare professionals, and to promote improved case management, including onward referral to SHCs for STI and HIV testing.[15] The campaign was led by PHE in collaboration with THT through social media, the gay press and leaflets and posters displayed in SHCs. Given the sustained increase in cases, PHE repeated the awareness campaign in 2016 in collaboration with THT, the LGBT (Lesbian, Gay, Bisexual and Trans) Foundation and Do it London (a LA initiative running city-wide sexual health promotion campaigns).[157] However, the follow-up evaluation conducted at three SHCs in London found that overall awareness of shigellosis among MSM remained low (29%).[158]

2.7.4.1.3 Treatment and antimicrobial resistance

Shigellosis can often be managed conservatively. However, where indicated, the primary treatment for uncomplicated infection is ciprofloxacin with alternative therapies including azithromycin and ceftriaxone.[43] AMR to front-line treatments is well reported in *Shigella* spp. and can vary by country of acquisition and route of transmission.[22,45,159-161] Due to increasing global reports of ciprofloxacin resistance, *Shigella* spp. were included on the 2017 World Health Organisation list of pathogens that require the urgent development of new antimicrobials.[162] Reduced susceptibility to ciprofloxacin and/or azithromycin in isolates from MSM has been reported widely across North America, Australia, Europe and Taiwan.[11,22,23,145,160,163-167]

In recent years, WGS has been used to describe both the global and regional spread of different *Shigella* species, and sub-lineages of those species, among large networks of MSM.[22,168,169] WGS of clinical isolates of *S. flexneri* serotype 3a (1995-2014) identified a new lineage that had spread rapidly through Europe (including England), North America, and Australia via sexual transmission in MSM, and was distinct from lineages originating in Africa and Asia.[22] This lineage had acquired multiple AMR determinants, and successful sub-lineages (i.e. those which had expanded rapidly and were circulating most recently) had acquired high-level resistance to azithromycin, conferred by a large plasmid (pKSR100) carrying the AMR genes *mphA* and *ermB*.[22] Horizontal transfer of this plasmid among different *Shigella* species and sub-lineages of those species, may have facilitated epidemics of *S. flexneri* serotype 2a and *S. sonnei* sub-lineages among MSM in the UK, demonstrated by the rapid increase in cases following the introduction of the plasmid into those sub-lineages.[24] A highly-related plasmid has also been detected in MSM-associated sub-lineages of *S. flexneri* serotype 2a and *S. sonnei* circulating in Australia.[169] In 2015, a cluster of multi-drug resistant (including resistance to azithromycin and ceftriaxone), extended-spectrum beta-lactamase (ESBL)-producing *S. sonnei* was reported in England (n=9).[23] Resistance was conferred by the acquisition of the pKSR100 plasmid, but with an additional mobile genetic element that enabled ESBL production. Although these isolates were sensitive to ciprofloxacin, the potential for spread of this plasmid to other enteric pathogens is of concern, and it raises questions about future treatment options for shigellosis and the need for susceptibility testing.[23] Although

not the primary treatment for shigellosis, azithromycin is used for the treatment of several bacterial STIs, and it is hypothesised that resistance in MSM-associated *Shigella* spp. could be linked to off-target effects from high levels of antimicrobial exposure for STIs.[22,24] This is because MSM are disproportionally affected by STIs, including gonorrhoea, syphilis and LGV, resulting in high levels of antimicrobial exposure in this population.[96]

2.7.4.2 Hepatitis A

Hepatitis A is an acute and usually self-limiting infection caused by the Hepatitis A virus (HAV). Symptoms are more common in adults than in children and can include fever, malaise, nausea, anorexia and abdominal pain followed by jaundice, with severe morbidity and mortality more likely in those older than 50 years of age. Some people may experience relapsing hepatitis and in rare cases, acute liver failure.[170] Outbreaks associated with sexual transmission in MSM are well described.[171,172] A highly effective vaccine is available and in England (and other low endemicity settings), selective vaccination of individuals at higher risk is recommended, for example travellers to endemic areas and MSM. Active vaccination and/or passive immunoglobulin therapy is available for the management of contacts of HAV cases and for outbreak control.[172]

2.7.4.2.1 Recent epidemiology

In December 2016, the European Centre for Disease Prevention and Control (ECDC) reported an increase in hepatitis A cases predominantly affecting MSM in Europe. Germany, the Netherlands and the UK were among the first countries in the European Union (EU) to report an increase in cases.[173] In October 2016, a strain (RIVM_HAV16_090) of genotype 1A was identified in two MSM in the Netherlands associated with sexual contact at the 2016 EuroPride festival in Amsterdam - the viruses isolated from these men were genetically indistinguishable.[9] The infecting strain had not been seen in the Netherlands since 2010, and was genetically related to strains from Japan and most likely originated in Asia; it was found to be genetically indistinguishable (based on phylogenetic analysis of sequences from the VP1/2A region of the genome, the standard protocol for HAV strain discrimination) to the strain associated with a large outbreak in MSM in Taiwan that began in 2015.[9] In December 2016, the UK reported an increase in cases, mostly in MSM, with a different strain (VRD_521_2016)

of genotype 1A, which had not been previously reported in the UK and was related to strains from Latin America.[173] Subsequent phylogenetic and epidemiological analyses revealed likely importation from Spain, followed by secondary sexual transmission among MSM in the UK.[8] In January 2017, Germany reported three distinct clusters of genotype 1A in Berlin primarily among MSM: two of the clusters involved strains RIVM_HAV16_090 and VRD_521_2016 described above, while the third involved a new strain (V16_25801) identified in Berlin and other German cities.[10,16]

The three distinct clusters quickly spread across Europe (22 EU countries reporting 3,813 confirmed cases by December 2017) and beyond, reaffirming that the sexual networks of MSM in Europe are highly interconnected.[174] Two of the three strains associated with the European outbreaks were also likely imported to Israel by men who had travelled abroad.[17] Israel introduced universal vaccination for toddlers in 1999, which has led to a low incidence of infection in the general population but most adult men remain susceptible.[17] Increases in the number of HAV cases among MSM were also reported in Chile, Australia and New York City.[175-177] In Chile, cases peaked in mid-2017 representing a 168% increase compared to the same period in 2016, with the Santiago Metropolitan Region seeing the largest increase in cases.[18,175] Limited phylogenetic analysis for a small group of men found that the infecting strain was genotype 1A, and was related to one of the three clusters associated with the European outbreaks.[18,175]

The characteristics of the cases involved in these international outbreaks have been broadly similar. In the Chilean and Taiwanese outbreaks, the majority were MSM and a high proportion were living with HIV and had a history of, and/or coinfection with STIs or *Shigella* spp.[19,175] Over a third of cases in the UK outbreak reported one or more higher risk sexual behaviours including anonymous sex, using on-site sex venues, or the use of apps to meet partners.[178] Most cases were MSM living in London, but as the outbreak progressed, cases were reported in the wider population highlighting the spread of infection to contacts who were not sexual partners.[178] Enhanced surveillance data from seven EU countries suggested that among cases who travelled, a high proportion of cases reported sexual contact (67%).[179]

2.7.4.2.2 Public health response

Control measures focussed on vaccination of at-risk MSM and close contacts, as well as implementing enhanced surveillance and raising awareness among health professionals of the need to consider and test for HAV, and raising awareness among MSM in the community about the risk of sexual transmission.[8,9,178] In some regions, public health action was also required in the wider community, for example, mass vaccination of 1800 school children and teaching staff in the UK.[178] In Taiwan, the provision of free vaccination to people living with HIV or those under 40 years of age recently diagnosed with syphilis or gonorrhoea, coincided with reductions in cases at the end of 2016.[19,180] Vaccination of MSM living in outbreak areas or attending Pride events where riskier sexual contact was considered likely was recommended by ECDC during the European outbreaks in 2017, but implementation was hampered by a worldwide shortage of vaccine.[16]

Prior to this outbreak, BASHH guidelines recommended that at-risk MSM were offered vaccination when attending SHCs where increased rates of infection were reported locally.[181] In England, vaccination is commissioned at the LA level, and many clinics stopped offering it, particularly in areas where it was not considered cost-effective or affordable. This outbreak highlights the health risks that may be associated with such policy decisions. During the outbreak, PHE conducted public health campaigns alongside sexual health charities to raise awareness and promote vaccination. However, due to the global shortage of vaccine and a lack of consensus regarding funding, there were delays in initiating control measures. Ultimately, a central stockpile of vaccinations was procured by PHE for distribution to SHCs. This included temporary off-label use of paediatric vaccines to maximise the number of people who could be immunised.[178] The associated healthcare costs of the outbreak were estimated to be £1.5 million, primarily due to the high number of cases admitted to hospital.[178] The guidelines (last updated in 2017) now recommend that all MSM attending SHCs are opportunistically offered vaccination.[182]

The HAV outbreaks highlight the importance of international sexual networks in fuelling the spread of sexually transmissible infections. It has been suggested that sustained transmission

of HAV could be prevented if the level of immunity in the at-risk MSM population exceeds 70%.[183]

2.7.4.3 *Entamoeba histolytica*

E. histolytica is a protozoan parasite endemic in areas with poor water and sanitation infrastructure.[53] In the vast majority of cases, infection is self-limiting and asymptomatic, but in some cases can lead to invasive disease with amoebic dysentery or liver abscess.[53] There has been renewed interest in the transmission of *E. histolytica* as a sexually transmissible infection in non-endemic countries following reports of symptomatic amoebiasis among MSM in Japan, Taiwan, the Republic of Korea and Australia, mostly among people living with HIV.[184-187] In a cross-sectional study carried out among individuals attending an HIV voluntary counselling and testing service in Taiwan, the detection of antibodies for *E. histolytica* was associated with sex between men, oral-anal contact, HIV infection, older age and current syphilis.[188] The geographical distribution of this re-emerging infection in MSM is thought to reflect the higher background prevalence of infection in Asia.[184]

Two small clusters of invasive amoebiasis among MSM were reported in Barcelona, Spain in October 2016 and January 2017.[14] All cases reported oral-anal sex and some reported multiple sexual partners (up to 30 in the three months prior to infection) and attendance at sex parties. There was no epidemiological link found between the cases, suggesting that substantial under-diagnosis was likely. Half of the men were HIV-diagnosed (CD4 cell count >500 cells/mm3) and all but one had a concurrent or recent STI and/or had a concurrent *S. flexneri* infection.[14] Of note, at least one asymptomatic contact tested positive for *E. histolytica* while another was previously diagnosed with amoebiasis after travelling to Brazil, highlighting the spectrum of clinical illness and potential for international spread.

2.7.4.4 *Campylobacter* spp.

Campylobacter spp. are a major cause of bacterial gastroenteritis globally.[189] Extra-gastrointestinal infection is rare but can result in complications including Guillain-Barré syndrome (a neurological condition), bacteraemia, lung infection, meningitis or reactive arthritis, particularly in those who are immunocompromised or in the elderly.[189-191] The most

frequently isolated species is *C. jejuni,* followed by *C. coli*.[189,190] Other species are less common but have been associated with invasive disease in those who are immunocompromised or who have other co-morbidities (e.g. *C. fetus*).[192] Antimicrobial treatment is not required in most cases, but severe infections are treated with macrolides (clarithromycin, azithromycin or erythromycin) or ciprofloxacin.[11,193]

Several outbreaks of campylobacteriosis caused by infection with *C. jejuni,*[11,194] *C. coli*,[12,195] or *C. fetus*[136] have been reported among MSM in Canada over the past decade with most isolates exhibiting AMR to at least two of erythromycin, ciprofloxacin or tetracycline.[11,12,194,195] These outbreaks have predominantly occurred among MSM living in Montreal's "Gay Village" or surrounding areas. Sexual transmission of two multi-drug resistant (MDR) strains of *C. jejuni* resulted in a persistent outbreak in MSM lasting over 10 years.[11] The authors noted that MDR in domestically acquired *C. jejuni* is rare but has previously been reported in several MSM with AIDS in Australia.[196] As with other enteric pathogens, most cases in the Canadian clusters were MSM living with HIV who had history of, or were co-infected with STIs (syphilis, gonorrhoea or chlamydia) and/or other enteric pathogens (*Shigella* spp., *G. lamblia* or *E. histolytica*).[11,12,136,194,195] Among those diagnosed with HIV a range of CD4 cell counts (10 to over 1000 cells/mm^3) and HIV viral loads (<50 to >800,000 copies per ml) were reported, indicating that campylobacteriosis occurs in MSM who are both immunocompetent and immunosuppressed. MSM who were part of the cluster infected with *C. fetus*, a less well characterised species of *Campylobacter*, reported attending sex venues and the use of the internet to meet casual partners prior to symptom onset.[136]

2.7.4.5 Other enteric pathogens

Other causes of sexually transmissible enteritis include *G. lamblia*, *Salmonella* spp. and *E. coli*,[103,104] although there have been few reports in recent years. In the pre-HAART era, *Cryptosporidum* spp. and *Microsporidum* spp. were protozoan parasites commonly seen among MSM living with HIV, but outbreaks associated with sexual transmission are not commonly reported; infections are usually opportunistic given immunosuppression rather than a function of behaviour.[104] Cytomegalovirus was also an important diarrhoeal agent in the pre-HAART era but is now rarely observed.[104,197]

In December 2013, a small cluster of STEC O117:H7 serotype was detected among MSM in England (n=9).[13] Most men were living with HIV and reported multiple sexual partners (median of five partners in the two weeks before symptom onset) and engaging in higher risk behaviours such as chemsex and fisting. STEC O117:H7 is rare in England with just 13 cases reported between January 2009 and November 2013, with most associated with travel to tropical destinations.[13,198] Genomic analyses suggested likely importation from Latin America, and evidence of horizontal transfer between the outbreak strain and MSM-associated *Shigella* spp. circulating in England during the same time-period; the outbreak lineage was also resistant to azithromycin.[199]

2.7.5 Surveillance of enteric pathogens in MSM in England

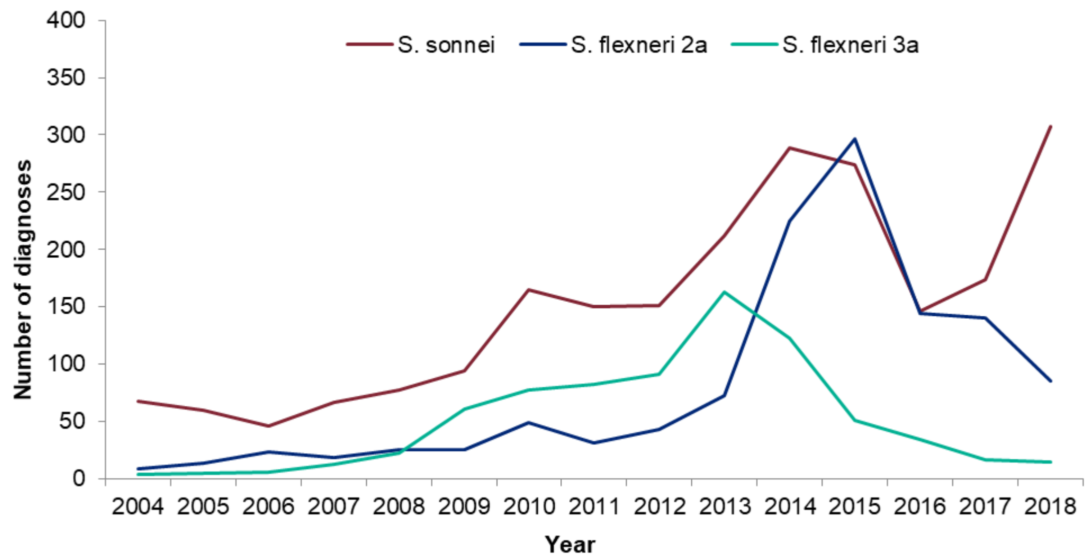
In England, laboratory reporting data form an integral part of national surveillance activities for enteric pathogens. These data provide a timely source of information for investigating trends and identifying potential outbreaks. In addition, they were essential for detecting and monitoring recent national outbreaks of shigellosis and hepatitis A associated with sexual transmission between men, as described in sections 2.7.4.1 and 2.7.4.2 above.

2.7.5.1 Shigella spp*. surveillance*

In England, routine surveillance of enteric pathogens in MSM has primarily focussed on *Shigella* spp. using data generated by PHE following specimen referral to the GBRU. The epidemiology has changed remarkably in the last 10 years. Traditionally, most cases were associated with foreign travel. However, since 2004 the number and proportion of UK-acquired cases without a reported foreign travel history has increased and in 2010, started to exceed the number of travel-associated cases.[200] To date, an excess in adult male cases (≥16 years old) with no recent foreign travel has been used to infer sexual transmission in MSM since routine laboratory reports do not contain information on sexual identity or behaviour.[5] Since 2009, the data have shown successive waves of different *S. flexneri* serotypes and *S. sonnei* in adult men with no foreign travel history, while laboratory reports in adult women have remained relatively stable (Figure 2.3). These epidemiological trends, combined with data on the male-to-female (M:F) ratio are consistent with sexual transmission of *Shigella* spp. through

sex between men.[5,201] In 2014, at the height of the shigellosis epidemic, the M:F ratio in adults with no foreign travel reached a maximum of 30.8 for *S. flexneri* serotype 3a and 16.1 for *S. flexneri* serotype 2a and the percentage of cases that were men was 96.9% and 94.1%, respectively. The M:F ratio for *S. sonnei* reached a maximum of 2.9 in 2015, and the percentage of cases that were men in the same year was 73.9%.

a) Men



b) Women

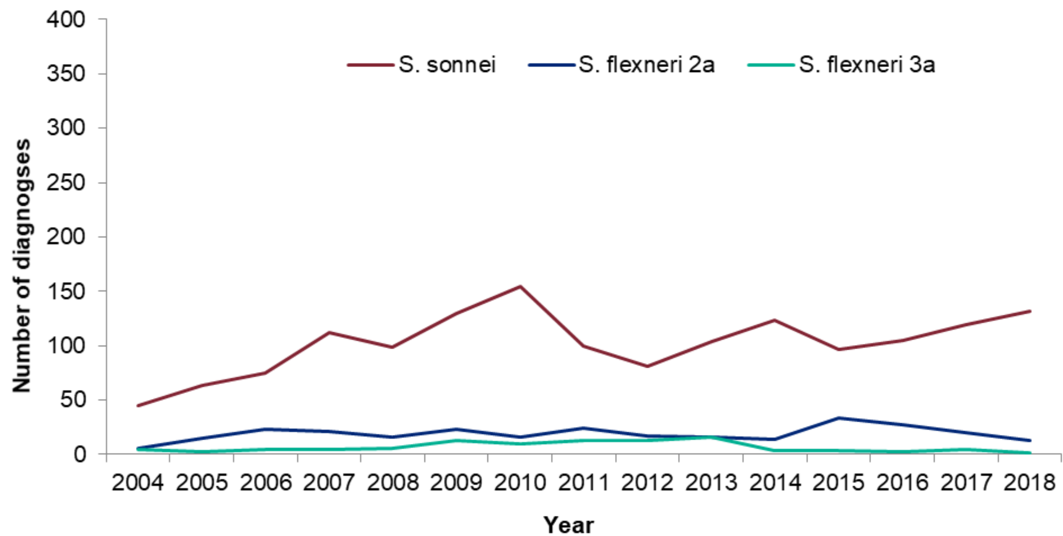


Figure 2.3: Non-travel associated diagnoses of *S. flexneri* serotype 2a, *S. flexneri* serotype 3a and *S. sonnei* in England (≥16 years old), by year, 2004-2018

Data from: Public Health England, Gastro Data Warehouse (GDW)

Between 2015 and 2016, total diagnoses of *S. flexneri* serotype 2a, *S. flexneri* serotype 3a and *S. sonnei* in adult men with no foreign travel history fell by nearly half from 622 to 324 (47.9% decrease). The reasons for the sharp recent decline in diagnoses during this time are unclear but may reflect a real drop in transmission, due to modified sexual behaviour or

changes to the size of the susceptible population and levels of population immunity (see section 2.7.4.1.1). The drop could also reflect an artefact of laboratory surveillance data. For example, changes in local reporting practices by primary diagnostic laboratories may have influenced the number of samples sent to the national reference laboratory for species identification and typing. In addition, there could have been a reduction in the number of people seeking healthcare or in the number of samples collected for microbiological investigation. It seems unlikely that these factors would account for the decrease in the number of *Shigella* spp. diagnoses as there have been no changes to guidelines for sample collection or reporting. In 2018, there was a resurgence in diagnoses of *S. sonnei* in adult men with no foreign travel history suggesting that transmission in sexual networks of MSM persists (Figure 2.3). These data suggest that current health promotion messages and awareness campaigns targeting MSM have been unable to control the spread of *Shigella* spp.

2.7.5.2 Use of gender distribution to detect enteric pathogens acquired through sex between men

Over an 11-year period (January 2003 to December 2013), Mook *et al.* (2018), explored use of the M:F ratio and the total proportion of enteric infection diagnoses that were in men among people with no foreign travel history to identify cases potentially associated with sex between men.[201] An excess of adult male cases was observed for laboratory reports of *Shigella* spp., *E. histolytica*, *S.* Typhi, *G. lamblia*, *Campylobacter* spp. and HAV, suggesting that sexual exposure of some enteric pathogens may be unidentified (Table 2.6). Further stratification by geographical region and age group revealed stronger signals in regions with higher STI rates and larger MSM populations such as London, Brighton and Manchester (referred to as 'high-risk regions'). For example, *Shigella* spp. showed the strongest signals in people aged 25-49 years and in those living in 'high-risk' regions from 2010 onwards, consistent with the known *Shigella* spp. outbreaks.[5,201] In addition to *Shigella* spp., male excess signals were strongest for *E. histolytica* and HAV, where an excess was detected across multiple years for adults aged 25-49 years, and for *E. histolytica*, an excess was frequently detected in 'high-risk' regions.[201]

Other countries have used a similar approach. M:F ratios for 10-49 year olds helped identify potential MSM transmission for a range of STIs and enteric infections, and a hepatitis A outbreak in MSM, in the state of Georgia, USA during 1998-2001 and highlighted a potential increase in shigellosis and giardiasis among MSM.[202] In British Columbia, incidence rates for shigellosis and amoebiasis were highest in adult men (20-59 years) and provided evidence of transmission associated with sex between men.[203]

The M:F ratio or the percentage of cases that are adult men can be useful indicators of potential sexual transmission of enteric pathogens in MSM but there are several limitations of these approaches. They assume a natural baseline gender ratio of 1:1, however, disparities in the gender distribution may reflect differing patterns of health-seeking behaviour, testing and/or reporting, and travel, food consumption or childcare practices.[201] There may also be increased transmission in other population groups disproportionately represented by men aged 25 to 34 years, such as people who inject drugs[204,205] or homeless people.[204,206] In addition, the M:F ratio may be sensitive to the absolute number of cases reported. For example, pathogens which are common in the general population, such as *Campylobacter* spp., may have a M:F ratio just above one, but the absolute number of excess adult male cases may still be considerable (Table 2.6).[201]

Table 2.6: Excess number of male cases, male-to-female ratio and percentage male by pathogen, risk area and age group, for laboratory-confirmed gastrointestinal infections with no reported travel history, England, 2003-2013

Organism	Total cases 16–65 years	No. excess males 16–65 years	Male-to-female ratio in 16–65 years	% Male			
				16–65 years all areas	16–65 years in high-risk areas	16–65 years in low-risk areas	<16 years
Campylobacter spp.	382 641	21 649	1.12	52.8 (52.7–53.0)	52.1 (51.6–52.6)	52.9 (52.8–53.1)	60.3 (60.0–60.7)
Cryptosporidium spp.	17 907	−4643	0.59	37.0 (36.3–37.7)	45.8 (43.1–48.6)	36.4 (35.6–37.1)	56.5 (55.9–57.2)
Entamoeba histolytica	1581	579	2.16	68.3 (65.9–70.6)	74.4 (70.9–77.6)	63.5 (60.3–66.7)	58.4 (48.8–67.6)
Giardia spp.	24 776	2396	1.21	54.8 (54.2–55.5)	58.5 (56.9–60.1)	54.2 (53.5–54.9)	58.2 (56.9–59.4)
Hepatitis A	3564	792	1.57	61.1 (59.5–62.7)	54.0 (51.0–57.1)	64.0 (62.1–65.9)	51.8 (48.3–55.2)
Norovirus	12 638	−858	0.87	46.6 (45.7–47.5)	52.5 (49.0–56.0)	46.2 (45.3–47.1)	56.0 (54.9–57.2)
Salmonella spp. (non-typhoidal)	45 910	−816	0.97	49.1 (48.7–49.6)	48.8 (47.8–49.8)	49.2 (48.7–49.7)	53.1 (52.5–53.8)
Salmonella spp. (typhoidal)	1483	161	1.24	55.4 (52.9–58.0)	56.4 (52.8–59.9)	54.4 (50.6–58.1)	53.6 (48.9–58.3)
Shigella spp.	10 506	272	1.05	51.3 (50.3–52.3)	62.5 (60.9–64.0)	44.8 (43.6–46.0)	52.1 (50.2–54.0)
S. flexneri	2753	651	1.62	61.8 (60.0–63.6)	73.7 (71.1–76.3)	53.2 (50.7–55.7)	50.2 (46.8–53.6)
S. sonnei	6171	−261	0.92	47.9 (46.6–49.1)	60.0 (57.9–62.2)	41.9 (40.4–43.5)	52.7 (50.1–55.3)
S. flexneri PT2a	234	102	2.55	71.8 (65.6–77.5)	86.4 (78.5–92.2)	58.9 (49.7–67.6)	44.0 (33.2–55.3)
S. flexneri PT3a	338	266	8.39	89.3 (85.6–92.4)	99.0 (96.3–99.9)	85.9 (78.7–91.4)	34.2 (19.6–51.4)
VTEC	3117	−673	0.64	39.2 (37.5–40.9)	35.5 (30.0–41.3)	39.6 (37.8–41.4)	51.4 (49.8–53.1)
Total	504 123	18 859					

Source: Mook *et al.* (2018).[201] Ratios above a threshold of two or where the percentage male has a lower confidence interval above 50% are shaded.

2.7.6 Prevalence of enteric pathogens in MSM

Few studies have explored the prevalence of enteric pathogens in MSM and comparing the results from these studies requires careful consideration of the study population, the proportion who had gastrointestinal symptoms and the diagnostic methods used. Most studies were carried out when many men living with HIV were likely immunosuppressed. Nonetheless, most studies suggest the prevalence of enteric pathogens in some populations of MSM may be high.

In the late 1970s and early 1980s, several cross-sectional studies investigated the prevalence of enteric protozoa, including *G. lamblia* and *E. histolytica/dispar* complex, among MSM in urban areas, predominantly in North America. These small studies recruited MSM from either clinic[93,94,207-209] or community-based settings[93,95] and participants were requested to provide stool specimens for microscopic examination and to complete a questionnaire. Prevalence estimates ranged from 20-36% for *E. histolytica/dispar* complex and 4-18% for *G. lamblia*, and non-pathogenic protozoa were often detected at a high frequency. In some cases, there was no association between the presence of gastrointestinal symptoms and a protozoan pathogen,[93,94,207,209] indicating the potential role of asymptomatic carriage in MSM populations. In these early studies, it is likely that gastrointestinal symptoms were caused by other untested enteric or rectal pathogens. Furthermore, some study populations were more likely to include symptomatic men, which could have resulted in biased estimates; the proportion with gastrointestinal symptoms ranged from 32% to 77%. Some studies found higher prevalence of enteric pathogens in MSM compared to other population groups.[93-95] For example, Phillips *et al*. (1981) used a non-selective strategy to recruit 163 men (51 gay, 48 bisexual, 64 heterosexual) and 17 women from a SHC in New York.[94] The prevalence of *E. histolytica/dispar* complex was 19.6% in gay men, 2.1% in bisexual men and 0% in heterosexual men and women, and for *G. lamblia* the prevalence was 3.9% for gay men, 4.2% for bisexual men, and 0% for heterosexual men and women. Markell *et al*. (1984) recruited 508 MSM through community-based venues in the San Francisco Bay Area and found the prevalence of *E. histolytica/dispar* and *G. lamblia* to be 28.5% and 5.7%, respectively.[95] In comparison, among 415 people who were representative of the general population within the

San Francisco Bay Area, *E. histolytica/dispar* and *G. lamblia* were detected in only 0.7% and 1.7% of samples, respectively.[210] Importantly, none of these early studies distinguished between the pathogenic *E. histolytica* and non-pathogenic *E. dispar* as suitable techniques were not widely available. Thus, the high prevalence was likely explained, at least in part, by commensal protozoans.[211,212] This was reflected in a UK study that detected *E. histolytica/dispar* complex in 20% of 225 MSM and 0 of 129 heterosexual men attending a London SHC; electrophoresis on 75% of the isolates found they were all non-pathogenic *E. dispar*.[211] Nonetheless, these studies suggested a high level of faecal exposure associated with sexual contact in some MSM populations.

The prevalence of BEPs was also investigated among MSM attending SHCs in the 1980s. In 119 symptomatic (enteritis, proctocolitis or colitis) and 75 asymptomatic MSM attending a SHC in Seattle in 1983, the prevalence of *Campylobacter* spp. was 7% and 3%, respectively, and for *S. flexneri* the prevalence was 3% and 1%, respectively.[213] Where tested, *E. histolytica/dispar* complex was detected in 29% (20/70) of symptomatic MSM and 25% (6/24) of asymptomatic MSM, and *G. lamblia* was detected in 14% and 4%, respectively.[213] In the late 1980s, Laughon *et al*. (1988) estimated the prevalence of enteric pathogens and rectal STIs in MSM participating in an HIV cohort study in Baltimore, USA.[214] In 243 asymptomatic men, 6.4% (*Campylobacter* spp. 2.6%, *G. lamblia* 2.1%, *E. histolytica/dispar* complex 1.2% and *Shigella* spp. 0.4%) had an enteric pathogen detected. Although none of these men had AIDS at recruitment, the detection of an enteric pathogen or rectal STI was higher among individuals who were living with HIV and had lower CD4 counts (<400 cells/mm^3). There were differences in the study population and the geographical location between these studies, however, the authors noted that changes in sexual behaviour following the onset of the HIV/AIDS epidemic may have resulted in a reduction in prevalence over time.[214]

Studies providing more recent prevalence estimates reflecting contextual changes in HIV treatment and sexual behaviour are limited by small sample sizes and convenience sample study populations, with risk of bias. A study of 175 HIV-negative MSM (23% had a history of diarrhoea in the past three months) attending a SHC in Edinburgh during the mid-1990s detected *E. histolytica/dispar* complex in 9% of stool specimens and *G. lamblia* in 3%.

However, these data should be interpreted with caution because there were significant behavioural differences between patients who provided a stool sample and those who did not.[215] In Italy, *G. lamblia* was detected in nearly 17% of 74 asymptomatic MSM recruited through gay venues in Western Sicily in 2010.[216] A UK-based cross-sectional study estimated the prevalence of BEPs among a convenience sample of MSM diagnosed with rectal chlamydia during 2012 (n=444).[28] Residual rectal swabs from MSM attending 12 SHCs were tested using real-time PCR for a range of BEPs (*Shigella* spp., *Campylobacter* spp., enteroaggregative *E. coli* (EAEC) and *Salmonella* spp.) generating an overall prevalence estimate of 8.6% (95% CI: 6.3% to 11.6%): *Shigella* spp. were detected in 1.8% (95% CI: 0.9% to 3.6%), *Campylobacter* spp. in 1.8% (95% CI: 0.9% to 3.6%) and EAEC in 5.2% (95% CI: 3.5% to 7.7%). None of the specimens tested positive for *Salmonella* spp. or STEC. There was some evidence that prevalent BEPs were associated with symptom presentation (13.1% in symptomatic vs 6.4% in asymptomatic; p=0.05) and HIV (12.6% in HIV-diagnosed vs 6.3% in HIV-negative; p=0.05). The prevalence of *Shigella* spp. was higher in MSM who were living with HIV (4.7% in HIV-diagnosed vs 0.5% in HIV-negative; p=0.01). About half of specimens that tested positive were from asymptomatic cases suggesting that asymptomatic carriage may play a role in sustaining transmission of BEPs among MSM.[28] Importantly, the study used stored residual swabs that had previously had DNA extracted for chlamydia and LGV testing. This might have reduced the quantity and quality of DNA available, underestimating prevalence in the population.

2.7.7 Transmissibility of enteric pathogens in MSM

Understanding the dynamics of enteric pathogen transmission in MSM is important for interpreting epidemiological trends and informing the development of interventions. At a population level, the spread of a pathogen in a given population can be measured by the average number of new cases originating from each infected person. In epidemiology, this measure of transmissibility is known as the reproduction number and there are two forms, the basic reproduction number (R_0) and the case reproduction number (R).[217]

R_0 estimates the maximum potential for transmission when the pathogen is introduced into a totally susceptible population. For a STI, R_0 depends on i) the average probability of

transmission given contact between an infected and susceptible person (β), ii) the average number of new sexual partners (c), and iii) the average length of time a person is infectious for (D). Mathematically, this is expressed as $R_0=\beta cD$. When $R_0>1$, the infection will spread through a population.[217]

There are several factors that influence the R_0, including the biology of the pathogen, the propensity and duration of clinical symptoms, individual sexual behaviour and patterns of sexual mixing.[218-221] Heterogeneity in terms of sexual behaviour and partner change in MSM means that small sub-groups of the population who are highly sexually active can have a major influence on the spread STIs, and potentially enteric pathogens. In addition, understanding patterns of sexual mixing within and across different risk groups can help to explain the distribution and persistence of infection. Unlike most STIs, it is also important to consider that some enteric pathogens may induce long-term (e.g. HAV) or transient immunity (e.g. *Shigella* spp.) (see section 2.7.4.1.1). In such circumstances, the number of new cases originating from each infected person is estimated by the case reproduction number (R), which estimates actual transmission given that a proportion of the population is immune or already infected.[217]

Any intervention that aims to reduce the prevalence or incidence of infection should influence the components that determine R_0.[219] For example, the promotion of hygiene measures before and after sex, and the use of barrier methods for oral-anal sex, attempt to reduce the probability of enteric pathogen transmission between an infected and susceptible MSM (β).

2.7.8 Summary and evidence gaps

Enteric pathogens in MSM warrant public health attention due to the increasing frequency of reported outbreaks, rising trends in diagnoses, widespread geographic distribution and the development and spread of AMR. International sexual networks appear to facilitate transmission and introduce new strains into susceptible populations. Furthermore, the spread of AMR could have implications for the treatment of enteric and other pathogens.

My review has highlighted several important gaps in the literature where improved knowledge could help to inform the development and implementation of better control measures:

- Recent and representative data on the prevalence and distribution of different enteric pathogens in the MSM population are not available. Whilst enteric pathogens and their clinical syndromes are notifiable, the number and size of enteric pathogen outbreaks is likely to represent only a fraction of cases since many symptomatic individuals will not seek care or provide stool specimens for clinical diagnosis.[7,74,76] Furthermore, asymptomatic and/or persistent infections may play a role in facilitating the spread and maintenance of enteric pathogens in MSM populations.[7,28,222]
- The specific behavioural and contextual risk factors associated with transmission in MSM are not well described, particularly for anything other than *Shigella* spp. A better understanding of these factors could improve our understanding of the spread and persistence of enteric pathogens in MSM and help to identify people that may be at risk of acquiring an infection.
- Current data from outbreak investigations and national surveillance suggest there is an association between enteric pathogens, STIs and HIV, and that these infections occur in overlapping sexual networks. However, the underlying factors explaining the relationship between HIV and enteric pathogens remain unclear.
- Enteric pathogens in MSM are often resistant to antimicrobials, but the drivers of AMR and how they overlap with antimicrobial exposure for STI treatment requires further exploration and might inform guidelines on antimicrobial usage.
- Surveillance systems need to be sensitive and timely enough to identify potential sexual transmission in MSM, distinguish sexual from non-sexual transmission, and ensure that appropriate public health action is taken to control the spread of infection.

Chapter 3: Prevalence and risk factors of bacterial enteric pathogens in MSM: a cross-sectional study at a London sexual health clinic

In this chapter, I describe the design, implementation and analysis of a cross-sectional study of bacterial enteric pathogens (BEPs) among MSM attending the UK's largest sexual health clinic (SHC), Dean Street. I chose to focus on BEPs, which are a sub-set of all enteric pathogens that have been associated with recent outbreaks in MSM. The reasons for this were i) capacity to conduct the study within the constraints of a PhD timeframe, and ii) relative importance of BEPs in the MSM population, particularly successive epidemics of different *Shigella* spp.

3.1 Introduction and rationale

As discussed in the previous chapters, BEPs in MSM are a global public health concern due to the increasing number of reported outbreaks and the development of AMR. There are limited data on the burden of infection among MSM, particularly for pathogens other than the shigellae. To date, most data are from clinical case reports, outbreak investigations or laboratory surveillance data, which are all reliant on symptomatic individuals presenting for healthcare and these data almost certainly underestimate the true number of infections. We have very limited understanding about asymptomatic carriage of these pathogens, which might play an important role in sustaining transmission within specific sexual networks of MSM. There is a need to understand the underlying prevalence and risk factors of BEPs in MSM and their relationship with STIs and HIV so that improved public health and clinical control measures can be developed and implemented.

The specific research questions addressed in this chapter are as follows:

1. What is the overall prevalence of BEPs in MSM?
2. What is the role of asymptomatic carriage or subclinical infection in sustaining transmission of BEPs among MSM?

3. What is the relationship between BEPs and sexual risk behaviours in MSM and how does this overlap with STIs and HIV?
4. What is the prevalence of azithromycin resistance in BEPs in MSM and its relationship with previous treatment for STIs?

To help answer these questions, I conducted a cross-sectional study in a sample of MSM attending a large SHC in central London. MSM attending SHCs are unlikely to have severe symptoms of gastrointestinal illness but represent a sexually active population that are likely to be at increased risk of acquiring enteric infections through faecal-oral transmission linked to sexual activity, especially oral-anal contact. As such they are an appropriate population in which to explore the prevalence of, and risk factors for, these infections. Stool specimens are the recommended specimen for the clinical diagnosis of enteric infections, however, it is not usually practical to obtain these in SHCs, which would require patients to take away a stool collection kit and return a sample. A previous feasibility study showed that rectal swabs provide a practical and cheap alternative method of detecting BEPs for research purposes,[28] and I used this approach for my study.

3.2 Aim and objectives

The aim of the study was to better understand the epidemiology of BEPs in MSM to enable improved infection control. The objectives were to:

1. Provide up-to-date prevalence estimates for selected BEPs among MSM routinely attending a large SHC in central London
2. Determine the clinical, socio-demographic and behavioural risk factors associated with BEPs among MSM routinely attending a SHC
3. Explore the prevalence of azithromycin resistance and whether this is associated with a previous diagnosis of a bacterial STI

3.3 Methods

3.3.1 Study design

This was a cross-sectional study design at a single, large SHC in central London. All adult men who attended the clinic during the study period and who had a rectal swab taken for *Chlamydia trachomatis* and *Neisseria gonorrhoeae* testing were anonymously screened for selected BEPs using an opt-out approach (see section 3.3.5) and with all patient identifiers removed prior to testing (see section 3.3.9). The results obtained from BEP detection were linked to clinical, socio-demographic and behavioural data extracted from the clinic database and to the GUMCAD national STI surveillance system (see sections 3.3.7 and 3.3.8).[223]

When designing this study, I also considered an alternative cross-sectional study design that invited adult men attending the SHC to take part in a study (i.e. an opt-in approach with informed consent) involving completion of a self-administered study questionnaire and the collection of an additional rectal swab. I discussed the feasibility of the two different study options with several clinicians. The decision process underlying my choice of the opt-out approach using routine data and residual rectal swabs was guided by the following:

1. The need to obtain a study population that was reasonably representative of sexually active MSM attending urban SHCs in England.
2. The time needed to reach the required sample size and thereby complete the study within the time constraints of a PhD.
3. The capacity of the SHC to support recruitment and take informed consent.
4. The availability and type of behavioural data items routinely collected by the SHC that could be used in the study.

The main advantage of pursuing the opt-out approach was the increased likelihood of obtaining a large and reasonably representative study population within a shorter timeframe. I was concerned that the requirement to take informed consent (i.e. the opt-in approach) could have increased the likelihood of an unrepresentative sample, resulting in selection bias. This concern was supported by findings from previous cross-sectional studies conducted among

SHC attendees involving questionnaire completion, where response rates have ranged from 25% to 76%.[224-227] Low response rates often reflect the limited capacity of the SHC to implement the study alongside other research.[224,225,228] My discussions with clinicians at the SHC revealed that there was limited local capacity to support recruitment. As a result, there was an increased possibility that the characteristics of those who participated could be different from those who did not participate. Based on the expected number of weekly attendances at the SHC (see section 3.3.2), I estimated that the opt-out approach would take up to four weeks to reach the required sample size (see section 3.3.4). The recruitment period of the study would have been considerably longer using the opt-in approach and was not considered feasible within the time constraints of my PhD. This assumption was based on evidence from previous studies, including a cross-sectional self-administered questionnaire study conducted among MSM attending the same SHC included in my study; only 585 MSM were recruited over a 15-month period.[226]

The main disadvantage of the opt-out approach was that it restricted me to using only routinely collected data for risk factor analyses. Fortunately, the SHC in my study uses a standardised clinical proforma to collect comprehensive behavioural information. However, it was not possible to capture information outside of this proforma that could have been pertinent to BEP transmission such as specific sexual practices, recent foreign travel or gastrointestinal symptoms. In addition, missing responses or inaccurate reporting are key limitations of routine data collection, which can lead to bias. These issues are discussed further in section 3.5.2.

3.3.2 Study setting

The study took place at Dean Street (DS), the largest sexual health and HIV service in the UK and based in central London. The service consists of two clinics, Dean Street Express (DSE) and 56 Dean Street (56DS). The former is a sexual health screening service for people who are symptom-free but would like a check-up, and the latter is the main clinic for people with symptoms, those needing ongoing support or those requiring specialist services such as HIV post-exposure prophylaxis or HIV care. A service evaluation conducted during a one-year period from 2014 to 2015 found that approximately 75% of attendances across the service were to DSE and 25% to 56DS.[229]

At DSE, individuals first complete a clinical proforma on their sexual history using a touchscreen computer. They are then directed to a cubicle where they take self-collected swabs, which are delivered to the on-site laboratory for chlamydia and gonorrhoea testing (see section 3.3.6). Following this, a health advisor consultation takes place to review their sexual history and to take blood for syphilis, HIV and hepatitis B/C testing as appropriate. Test results are delivered by automated text message on the same day. At 56DS, all consultations are face-to-face with a clinician, including a full sexual history and collecting of samples as appropriate.[229] Sexual history is taken by the clinician using the clinical proforma, and the responses are subsequently entered onto the clinic database. If the patient is symptomatic, swabs will be taken by a clinician. All test results are delivered by automated text message within seven days.

DS provides sexual health and HIV services to a large number of MSM. When designing this study, I conducted a preliminary analysis using GUMCAD STI surveillance data to explore the context and population of the clinic. In 2016, there were approximately 2400 attendances per week across the two clinics, of which over half were by men who reported being gay or bisexual. In the same year, DS accounted for nearly one third (5228/17294) of all gonorrhoea diagnoses and one fifth (320/1557) of all new HIV diagnoses among MSM attending SHCs in England.

3.3.3 Study population

All MSM attending DS are routinely offered testing for *C. trachomatis* and *N. gonorrhoeae* from urine, pharyngeal and rectal swabs, regardless of symptoms. Although most rectal swabs are collected from men who identify as gay or bisexual, rectal swabs are also collected based on self-reported sexual behaviour and a smaller number are collected from other MSM, including heterosexual-identifying MSM, and from women.

3.3.3.1 Inclusion criteria

The study included all men aged 16 years or older, who attended DSE or 56DS during the study period and had a rectal swab collected for routine *C. trachomatis* and *N. gonorrhoeae* testing.

3.3.3.2 Exclusion criteria

The minimum age limit was 16 years. Any rectal swabs collected from people under 16 years old were not included in the study. In addition, any women with a rectal swab taken were not included in the study.

3.3.4 Sample size

As the primary objective was to estimate the prevalence of BEPs, I estimated the sample size to obtain precision ±1% of the estimated prevalence. To estimate a single proportion (e.g. prevalence of any BEP in the MSM population) I calculated a 95% confidence interval (CI) for a proportion (p) with a margin of error (d) using the following formula:

$$n = \frac{(1.96)^2\, p(1-p)}{d^2}$$

I estimated the sample size for two different outcomes, overall prevalence of any BEP and the prevalence of *Shigella* spp. In 2012, the estimated prevalence of BEPs among MSM diagnosed with rectal chlamydia at selected SHCs was 8.5% (95% CI: 6.3% to 11.6%).[28] The prevalence of *Shigella* spp. was 1.8% (95% CI: 0.9% to 3.6%). I used a more conservative estimate of prevalence to account for changes that may have occurred since 2012 and because the previous study used a convenience sample of men diagnosed with rectal chlamydia, who may have been at higher risk of enteric pathogens.

Assuming that overall BEP prevalence was 5% (p), a total sample size of 1825 men (approximately 91 positive specimens) was required to estimate prevalence in the sample to within 1% (d=0.01). This sample size was also sufficient to estimate the prevalence of *Shigella* spp. to within 0.5% (d) of 1% (p).

I hypothesised that the prevalence of an enteric pathogen was higher among men living with HIV compared to HIV-negative men. The formula below was used to calculate the sample size required to compare the prevalence between two groups:

$$n = \frac{\left(Z_{\frac{\alpha}{2}} + Z_{\beta}\right)^2 \ \pi_0(1-\pi_0) + \pi_1(1-\pi_1)}{(\pi_0 - \pi_1)^2}$$

Where n is the sample size of each group, $\pi 1$ and $\pi 0$ are the proportions in the two groups, $Z_{\frac{\alpha}{2}}$ represents the percentage point of the normal distribution corresponding to the significance level, Z_{β} represents the percentage point of the normal distribution corresponding to 100%-power.

A sample size of 1825 provided more than 90% power to detect a difference of 5% between two unequally sized sub-groups (assuming 1 in 5 men were living with HIV) at the 5% significance level if the prevalence among men living with HIV was 9% and the prevalence among HIV-negative men was 4% (assuming overall prevalence of 5%). If the outcome was less prevalent (i.e. overall prevalence of 1%), the sample size was sufficient to detect a difference of 2.5% if the prevalence among men living with HIV was 3% and the prevalence among HIV-negative men was 0.5%. The target sample size was increased by 20% to 2281 to account for missing data items (e.g. clinical or behavioural risk factors) and rectal swabs which failed DNA extraction or did not provide sufficient DNA for testing.

3.3.5 Opt-out approach

This study adopted an opt-out approach whereby posters and leaflets were displayed in the clinic waiting areas to inform patients about the study (Appendices 3.1 and 3.2). Patients could opt out if they preferred not to have their sample used in the study by signing their name against their clinic patient number in an opt-out log (Appendix 3.3). The clinic staff at DS were responsible for ensuring that these people signed the log, which was retained by the study research nurse at DS. The study research nurse generated a separate opt-out patient list, containing only the clinic patient number and date of attendance (i.e. no personal identifiers), to ensure that I could remove opt-out patients' specimens from the study (see section 3.3.8).

3.3.6 Specimen collection

All rectal swabs collected at DSE and 56DS are routinely processed on-site at DSE using the Cepheid GeneXpert CT/NG Assay. *C. trachomatis* positive swabs are subsequently sent to

North West London Pathology (NWLP) at Charing Cross Hospital for testing for lymphogranuloma venereum (LGV) *C. trachomatis* genotypes.

For this study, I collected residual rectal swabs (regardless of test result) from both the DSE laboratory and from NWLP (for the subset tested for LGV). All swabs were transported to the Gastrointestinal Bacterial Reference Unit (GBRU) at Public Health England (PHE) by courier following UN 3373 regulations for the transport of biological specimens (Category B). The swabs were transported in the original Cepheid Xpert tubes including residual transport buffer.

Each rectal swab was labelled with a barcoded clinic patient number and the date of attendance but had no other identifiable data (this is defined as pseudo-anonymisation by the Information Commissioner's Office).[230]

3.3.7 Epidemiological data collection

Clinical, socio-demographic and behavioural data were extracted from the GUMCAD national STI surveillance system[223] and the clinical database at DS.

GUMCAD is the national surveillance system for STIs in England and is managed by PHE. It is a pseudo-anonymised patient-level dataset that contains information about attendances, STI and HIV testing, and diagnoses made, at all SHCs in England.[223] GUMCAD contains clinic patient number, gender, age, sexual orientation, Lower Layer Super Output Area (LSOA) of residence, ethnicity and country of birth, but no patient identifiable data such as date of birth, name and postcode (Table 3.1). Patient records can be linked within, but not across clinics using the clinic patient number. A surveillance scientist within the HIV & STI Department at PHE extracted GUMCAD records for the study patients using the clinic patient number (see section 3.3.8), and grouped the data into categories, where appropriate, to minimise the risk of deductive patient identification (e.g. five-year age group was provided instead of age).

The clinic database at DS includes behavioural and clinical data items that are collected by the clinical proforma (see section 3.3.2) or as part of the patient consultation (56DS only). For this study, I included all data items that were collected by the clinical proforma (Table 3.2). This includes data items that are not available in GUMCAD such as information on the number

of sexual partners and the use of HIV pre-exposure prophylaxis in HIV-negative men (PrEP) (i.e. the use of antiretroviral medicines before sex to reduce the risk of acquiring HIV). There is also a composite question which asks MSM whether they are interested in a set of specific behavioural practices, indicating that they are likely to be at higher risk of STIs and HIV ("Are you into any of these?": Fisting, injecting, bare backing, chemsex). Herein, this question is defined by the term 'interest in specific high-risk practices.' I also included data on symptoms of gastroenteritis (Table 3.2). It is assumed that people attending DSE do not have symptoms, however, a free text field is provided on the computer administered clinical proforma, and some people may choose to report symptom data here. All people attending 56DS are routinely asked about symptoms, including symptoms of gastroenteritis, and these details are entered within the free text field of the clinic database. Clinic staff responsible for patient care and study research nurses extracted clinical proforma records and symptom data from the clinic database at DS using the clinic patient number.

Table 3.1: GUMCAD surveillance system data items

Data item	Data item description
Attendance date	Date
SHHAPT code	STI surveillance code (Sexual Health and HIV Activity Property Type) This code provides information on HIV & STI diagnoses made and sexual health services provided.[231,232]
Gender	All men
Age group	5-year age group: 16-19, 20-24, 25-29,30-34,35-39,40-44,45-49, 50-54, 55-59, 60-64, 65-69, etc.
Sexual orientation	Heterosexual, homosexual, bisexual, not known
Ethnic Group	White, black Caribbean, black African, black other, mixed, Asian, other, not known
World Region of birth	UK, EU, Other Europe, Caribbean, sub-Saharan Africa, South Asia, Central America, North America, South America, other, not known
Local Authority of Residence	Local Authority of patient residence (326 in England)
Index of Multiple Deprivation±	5 quintiles generated (coded 1-5) by mapping Lower Layer Super Output Area (LSOA) of Residence
Attendance Type	New or follow-up episode

±Index of Multiple Deprivation (IMD) is an overall measure of relative deprivation for small areas in England.[233] It combines information on seven domains of deprivation: 1) income 2) employment 3) health, skills and training, 4) crime 5) barriers to housing and services 6) health and disability and 7) living environment. In this study, IMD was generated based on the Lower Level Super Output Area (LSOA) of Residence (small geographic areas designed to have a population size of between 1500 and 3000).[234]

Table 3.2: Clinical and behavioural data items extracted from the clinic database at Dean Street

Data item	Data item description	Data item format
1	Number of sexual partners in the past 3 months	Number
2	Number of new sexual partners in the past 3 months	Number
3	Last condomless sex	Within 72 hours/ 6 weeks/ over 6 weeks/ never/ not since last HIV test
4	Receptive anal sex in the past 3 months	Yes/ No/ Don't know
5	Receptive oral sex in the past 3 months	Yes/ No/ Don't know
6	Last HIV test	Never tested/ within last year/ more than a year-ago/ I am HIV positive
7	"Are you into any of these?": Fisting, injecting, bare backing*, chemsex§	Yes/No/Don't know
8	Are you currently using PrEP±	Yes/No/I am HIV positive
9	Syphilis diagnosis in past	Yes/No/Don't know
10	Clinic attended	DSE or 56DS
11	Any other information	Any information on gastrointestinal symptoms entered in a free text field

Data items 1-9 are based on standardised questions asked to all patients attending the DS service using a clinical proforma. All individuals attending 56DS are routinely asked about symptoms, including symptoms of gastroenteritis, and these details were extracted from the free text field of the clinic database. It is assumed that individuals attending DSE do not have symptoms, however, a free text field is provided on the computer administered questionnaire, and some patients may choose to report symptom data here.

Data item 7 is a composite question which asks men whether they are interested in a set of specific behavioural practices that could indicate they are at higher risk of STIs and HIV.

*Bare backing refers to condomless anal sex.

≠ Chemsex refers to sex while under the influence of drugs, often involving group sex

± PrEP refers to HIV pre-exposure prophylaxis. This is the use of antiretroviral medicines before sex to reduce the risk of acquiring HIV.

3.3.8 Data management, linkage and anonymisation

The data management, linkage and anonymisation algorithm for my study is presented in Figure 3.1 and summarised below:

- Each rectal swab was labelled with a barcoded clinic patient number and the date of attendance before being transported to the GBRU (step 1).
- Upon arrival at the GBRU and after removal of opt-out specimens, I anonymised the swabs with a unique study ID. At the same time, I generated an encrypted and password protected temporary electronic file (T1) that contained the study ID, the clinic patient number and the date of attendance. I sent this file to the surveillance scientist at PHE who retained it for the duration of the study to facilitate linkage to data from GUMCAD and the clinic database. I did not have access to T1 from this point onwards and it was stored securely on an encrypted drive at PHE, separately from other study files and the rectal swabs (step 2).
- The study research nurse generated a list of people who had a rectal swab collected during the study period but were not eligible for the study (i.e. women and men under 16 years of age). This list was sent to the surveillance scientist, who anonymised the list with the unique study ID and sent it to me so I could remove the relevant swabs prior to testing (step 3).
- I performed laboratory testing on all anonymised rectal swabs that were eligible for inclusion in the study (step 4).
- DS clinic staff extracted information from the clinic database using the clinic patient number. This file was sent to the surveillance scientist who anonymised the data with the unique study ID using T1 (step 5).
- The surveillance scientist extracted GUMCAD data and anonymised the records with the unique study ID using T1 (step 6).
- All anonymised epidemiological data (GUMCAD and DS clinic database) were sent to me. I linked these data to the biological test results using the unique study ID and attendance date (step 7).

Please see Appendix 3.4 for a detailed explanation of the procedures involved.

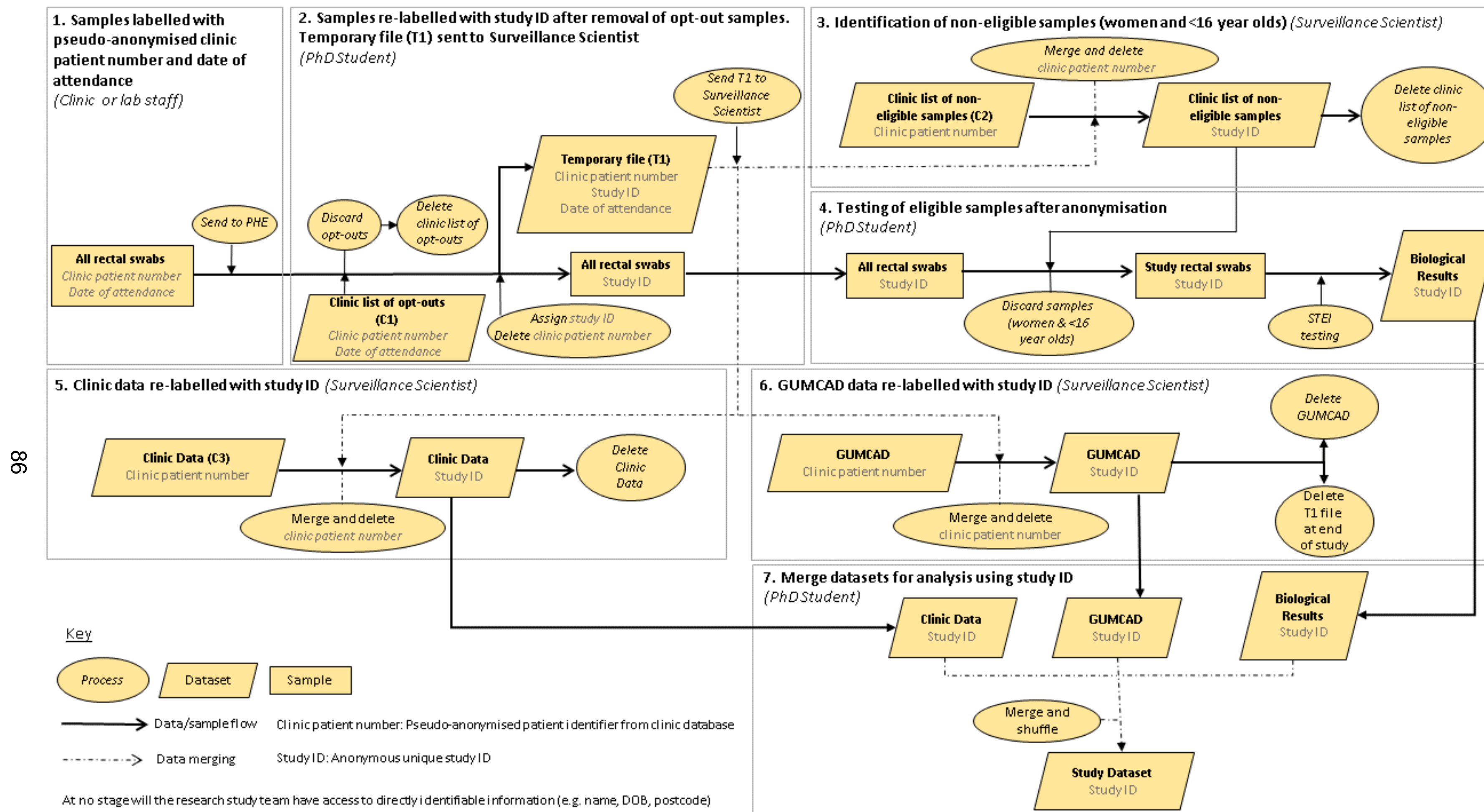


Figure 3.1: Data management, linkage and anonymisation algorithm for rectal swab specimens and epidemiological data collected at Dean Street

3.3.9 Ethical considerations

3.3.9.1 Consent

This study used an opt-out approach and did not request specific patient consent for the collection of rectal swabs used in this study. The reasons for this were as follows:

- The requirement to take consent may have increased the likelihood of an unrepresentative sample, whereby those taking part may have been different in their risk of BEPs. This could have undermined scientific rigour and biased the prevalence estimates.
- While I recognise the importance of patient autonomy in research, in this case, the study was very unlikely to lead to any harm (or benefit) to individual patients. The opt-out consent process provided patients with an opportunity to exclude their specimen from the study.
- The study did not impede the clinical pathway, impact on consultation time, or cause burden to NHS patients or staff, as the samples were already being collected as part of routine care.
- At no stage did I have access to personal identifiable data, including name, date of birth or postcode. As a result, there was no ethical requirement to request patient consent.

3.3.9.2 Anonymous testing without return of results

The use of rectal swabs, although useful for research purposes, has not been validated for clinical diagnostic use. In the absence of symptoms, positive test results for BEPs using this methodology have uncertain clinical implications. Use of anonymous testing ensures that test results cannot be returned and is appropriate where the clinical implications of the result are unclear. When an individual reserves a time-slot at DSE they are informed that they should have no symptoms. Individuals attending 56DS are asked about symptoms (including anal/rectal symptoms and symptoms of enteric infections) and these are managed according to national guidelines.[26]

3.3.9.3 Confidentiality

My study used pseudo-anonymised or anonymised rectal swabs and epidemiological data at all stages. The research study team outside of the local clinical/laboratory staff responsible for patient care did not have access to any patient identifiable data (e.g. name, date of birth or postcode). Clinical, socio-demographic and behavioural data (from GUMCAD and the DS clinic database) were anonymised with the unique study ID by a surveillance scientist, and the GUMCAD data were grouped into categories (where appropriate) to further minimise the risk of deductive patient identification (e.g. age group was provided instead of age). As a result, I was able to conduct the study without ever receiving the participants' identifiers. In addition, I performed BEP laboratory testing using anonymised rectal swabs and it was not possible for me to link the results to an individual patient. As an additional step, the surveillance scientist did not have access to the biological test results from BEP detection.

3.3.10 Ethical approval

The research protocol and all versions of the study documents (poster, leaflet and opt-out log) were approved by London Harrow NHS Research Ethics Committee (REC) and the NHS Health Research Authority (HRA) on 22nd November 2017 (REC ref 17/LO/1722, IRAS ID 225176) (Appendix 3.5). Chelsea and Westminster Hospital NHS Foundation Trust Research and Development (R&D) Office issued confirmation of Capacity and Capability on 29th November 2017 (Ref C&W17/056).

My PHE honorary contract was finalised in May 2017. This included all relevant data confidentiality agreements regarding access to national surveillance data (GUMCAD). This agreement also provided me with access to PHE IT equipment, laboratory materials and training. The PHE Caldicott guardian and the head of the PHE Research Support and Governance Office were made aware of, and advised on, the study procedures.

In England, the only way to access PrEP on the NHS is through the Impact Trial.[235] The PrEP Impact Trial is a large three-year trial that aims to answer outstanding public health and implementation questions that will facilitate effective introduction of a national PrEP programme in England. The PrEP Impact Trial started recruiting in October 2017, prior to the

start of my study. Given the sensitivity of the PrEP data, my study protocol was reviewed and approved by the PrEP Impact Trial Management Group.

3.3.11 Laboratory procedures

I carried out all laboratory procedures using standard operating procedures developed by the GBRU and described below. Training and guidance were provided by GBRU staff.

3.3.11.1 DNA extraction

The samples were gently agitated by pipette mixing and 400µl of the residual transport buffer was added to a single well of a sterile Thermo Scientific 96 square well storage plate (deep well). The samples were spiked with 10µl of modified green fluorescent protein (gfp) *E. coli*,[236] which acted as an internal positive control to minimise the risk of false negative reporting. The plates were processed as follows:

1. The plates were sealed and transferred to a bio-safe Centrifuge where they were spun at 3500 rpm for 20 minutes at 4°C.
2. The supernatant was removed using a pastette and the pellet re-suspended in 220µl of ATL buffer (tissue lysis buffer used for the purification of nucleic acids) by pipette-mixing, followed by 20µl of Proteinase K. The plates were re-sealed and incubated at 56°C in an Eppendorf ThermoMixer Block (shaking at 350rpm) for 45 minutes. This included 15 minutes to allow the lysate to reach the correct temperature.
3. 4µl of RNase A (Ribonuclease A) was added to the lysed cells and gently mixed. The plate was re-sealed and incubated at 40°C (shaking at 350 rpm) for 15 minutes. This step increases the purity of DNA.
4. The lysate was heat inactivated at 95°C for 20 minutes using a water bath. This step inactivated any remaining undigested organisms and enzymes.
5. The plates were transferred into the rotor of a bio-safe centrifuge and spun at 3500rpm for 20 seconds.
6. The sealing mat was removed and the plate transferred into the QIAsymphony machine. The DNA was extracted using the QIAsymphony DSP DNA Mini Kit (QIAGEN) and the elution volume was 100µl.

7. The eluted DNA was temporarily stored at 4°C followed by long-term storage at -20°C.

3.3.11.2 Real-time PCR detection methods

Eluted DNA extracted from residual swabs was used to detect a range of BEPs using real-time PCR primers and probes on a Rotor-Gene Q (QIAGEN). 2.5µl of extracted DNA was added to 22.5µl of reaction mix (forward and reverse primers at a concentration of 10µM each, the probe at a concentration of 1µM, Takyon™ MasterMix (Eurogentec) and nuclease-free water) to make a final reaction volume of 25µl. The amplification parameters were 95°C for 5 minutes, followed by 95°C for 15 seconds and 60°C for 60 seconds (40 cycles). The cycle threshold (CT) was set at 0.05 for all PCR targets. This method was based on the multiplex gastrointestinal PCR assay used by the GBRU for the detection of BEPs and includes gene targets for *Shigella* spp., *Campylobacter jejuni/coli*, STEC, EAEC, Enteropathogenic *E. coli* (EPEC) and *Salmonella* spp.[59] All PCR runs included negative (water) and positive (DNA known to contain the target gene) controls. CT values between 12 and 32 were considered a positive result, values between 33 and 35 were interpreted in combination with the curve result, while values >35 were considered a negative result.

In a secondary analysis, a real-time PCR method (developed and validated by GBRU) was used to detect the presence of *mphA*, an antimicrobial resistance gene associated with resistance to the macrolide azithromycin (S Nair, personal communication, June 2018).[237] Whole genome sequencing studies of MSM-associated *Shigella* spp. and STEC have described the presence of *mphA* on mobile genetic elements conferring resistance to azithromycin[22,168,169,199] and *mphA* has also been described in isolates of *E. coli* and *Salmonella* spp. (S Nair, personal communication, June 2018).[237,238] To enable assessment of the relationship between *mphA* detection and target BEPs, the *mphA* real-time PCR was performed on the Applied Biosystems TaqMan 7500 (Thermo Fisher Scientific) using all eluted DNA extracts which returned a positive result for one of the BEP target genes, and for comparison purposes, a random subset of 100 DNA extracts which returned a negative result for all BEP gene targets. I selected this number of samples based on what I considered feasible to process within the timeframe of the study. 2.5µl of extracted DNA was added to 22.5µl of reaction mix (forward and reverse primers at a concentration of 20µM each, the probe

at a concentration of 5µM, Takyon™ MasterMix [Eurogentec] and nuclease-free water). All PCR runs included negative (water) and positive (DNA known to contain *mphA*) controls. The amplification parameters were 95°C for 5 minutes, followed by 95°C for 3 seconds and 60°C for 30 seconds (40 cycles). The CT was set at 0.05 and CT values below 30 were considered a positive result. Details of all the primers, probes and gene targets are listed in Appendix 3.6.

3.3.12 Data analysis

3.3.12.1 Data cleaning and management

I cleaned and analysed all data using Stata v15. The distribution of variables was explored by tabulating and cross-tabulating the data. Missing values or inconsistencies in the data extracted from the clinic database were identified and sent to staff at DS to verify against the clinical records, for example, illogical or conflicting responses indicating both a HIV-negative and HIV-diagnosed status at the study attendance date. Any discrepant or missing responses that could be resolved were updated in the analysis dataset where appropriate.

All GUMCAD data are managed using standardised and validated cleaning algorithms at PHE. For example, codes denoting a specific STI diagnosis are de-duplicated to one per 42 days, reflecting the standard duration of an episode of care in SHCs.[223,239] Demographic data are also cleaned to remove conflicting responses (e.g. different ethnicities reported for the same patient over time). GUMCAD reporting does not include codes to indicate a negative test result. Therefore, according to recommended practice, I used the absence of relevant diagnosis codes for HIV or STIs to infer HIV-negative/unknown HIV status, and no/unknown previous STI diagnosis, respectively.[143,232,240] I generated a new variable to define the HIV status of each individual using data from both GUMCAD and the clinic database at DS.

Data on symptoms of gastroenteritis or diarrhoeal illness were cleaned differently according to whether a clinical proforma was completed (see section 3.3.2). For those who had a clinical proforma, the absence of data on symptoms of gastroenteritis or diarrhoeal illness was taken to indicate the absence of symptoms. For those who did not have a clinical proforma, symptom data were coded as missing.

3.3.12.2 Handling of missing data

The distribution of missing clinical, socio-demographic and behavioural data was assessed to understand the mechanism of missing data. This was important to understand the risk of bias and to guide the appropriate choice of methods for handling missing data.[241] Missing data can be classified as either missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR).[242,243] Data are said to be MCAR if the probability of missing data is the same for all cases and does not depend on the observed or unobserved data. MAR occurs when the missing data are dependent on observed data, but not on the missing data themselves. MNAR occurs when the probability of the missing data depends on the value of the missing value itself or on other unobserved data.[241,242] The pattern of missing data in the study can be used to indicate if the data are likely to be MCAR or not.[241,242] However, it is not possible to distinguish between MAR and MNAR.[241] In this analysis, differences between groups of individuals who were missing data and those who did not have missing data were assessed using the Chi-squared test and a p-value of <0.05 was considered statistically significant.

As described in section 3.4.3 below, it is plausible that the missing data in this study were MNAR. In such circumstances, there is no one method which can provide unbiased estimates.[244,245] For the primary results presented in this chapter, missing data were not imputed and available-case analyses were performed. However, several sensitivity analyses using simple imputation methods were conducted to assess the potential bias of missing data on the results. First, the missing indicator method was applied whereby missing data were grouped into an additional category, thus the full dataset was retained. Second, sensitivity analyses using worst-case and best-case scenarios were performed for all behavioural factors of interest and symptom data, where missing data were systematically replaced with the lowest or highest observed values to test the effect. Finally, single value imputation was used to assess the association between BEP detection and partner number by replacing missing values for the number of sexual partners with the median value. In this final sensitivity analysis, different categories of partner number were explored in addition to the inclusion of partner number as a continuous variable.

An additional sensitivity analysis using multiple imputation by chained equations (MICE) was also performed for the main analysis. This method uses the distribution of the observed data to estimate a set of plausible values for the missing data.[246]

I generated an imputation model for each variable that contained missing values (Appendix 3.7). Various model types were specified depending on the distribution of the variable, for example, logistic regression was used to impute binary variables. Each imputation model included the variables that were used in the analysis model, including the outcome. In addition, auxiliary variables that were both i) associated with the variable to be imputed, and ii) predicted whether that variable contained missing values were included in the model. 40 sets of plausible values were generated for each imputed variable, with 100 iterations (i.e. the number of cycles before the first set of imputed values was drawn). The estimates were combined to produce a single set of estimates using Rubin's Rules, which incorporates the uncertainty surrounding the missing data.[241,246]

After running the imputation models, the imputed values were compared to the observed values. In addition, convergence was checked for each variable by plotting the mean and standard deviation of the imputed values per iteration.

3.3.12.3 Descriptive and statistical analyses

Socio-demographic, behavioural, clinical and biological characteristics of the study population were tabulated to describe the study population. For continuous variables (e.g. number of new sexual partners in the past three months), the median and interquartile range (IQR) were calculated. The characteristics of the study population were explored in depth (e.g. association between reporting an 'interest in specific high-risk practices' and HIV status) using standard statistical techniques including Pearson's Chi-squared test (categorical or binary variables) and Wilcoxon rank sum tests (continuous and not normally distributed). A p-value of less than 0.05 was considered statistically significant.

To assess the representativeness of the study population, socio-demographic characteristics were compared to those of the wider population of MSM attending DS and all SHCs in England, using data collected from GUMCAD. Behavioural data were compared with data

available from four SHCs (two in London: Barnet Hospital, Croydon University Hospital, two outside London: Central Health Clinic University Hospitals Bristol, Southend University Hospital) collected during a pilot of the current GUMCAD system to collect behavioural and clinical data (2015 to 2016).[247] Data from Natsal-3, The third National Survey of Sexual Attitudes and Lifestyles (2010 to 2012),[248] were used to explore and compare the number of sexual partners reported in the past three months by MSM in the general population.

Prevalence estimates were calculated with 95% confidence intervals (CIs) using the Clopper-Pearson (exact binomial) method. The association between the detection of any BEP and clinical, socio-demographic and behavioural risk factors were explored using univariable and multivariable Poisson regression with robust error variances. This is an alternative to logistic regression in the analysis of cross-sectional data with a binary outcome and is used to directly estimate prevalence ratios (PR) with 95% CIs.[249] For multivariable analyses, each exposure variable (i.e. socio-demographic, clinical or behavioural factor) was adjusted in a separate model *a priori* for age (as a continuous variable), clinic (56DS or DSE) and HIV status. Overall p-values for heterogeneity were calculated using the Wald test and p-values for the test for linear trend were calculated for age group and number of sexual partners (new and total). I chose this modelling approach because many of the behavioural risk factors included in my analysis were highly correlated indicating that they do not occur independently from one another. The effect of HIV status on the relationship between each of the clinical, socio-demographic and behavioural risk factors was assessed for effect modification by including an interaction term in the multivariate models. The presence of an interaction would indicate the need to present the PRs stratified by HIV status.

In a secondary analysis, the prevalence of *mphA* was calculated with 95% confidence intervals (CIs) using the Clopper-Pearson (exact binomial) method. The association between *mphA* detection and a diagnosis of a bacterial STI in the past year was explored using Pearson's Chi-squared test. This study did not have any data on STI treatment. However, a diagnosis of a bacterial STI in the previous 12 months was used as a marker for previous antimicrobial exposure.

3.4 Results

3.4.1 Specimen collection

Specimen collection took place between 20th December 2017 and 6th February 2018. The flowchart in Figure 3.2 outlines the number of specimens processed, assessed for eligibility and tested for BEPs. 2507 specimens were received, of which 2399 were assessed for eligibility. No individuals opted out of the study. Of 2341 eligible specimens, 151 (6.5%) failed DNA extraction due to inadequate test results (i.e. the internal DNA extraction control failed) or due to technical errors with the QIASymphony machine. During the study, 74 repeat rectal swabs were received for 69 participants; one repeat swab was received for 64 participants and two repeat swabs were received for five participants. All subsequent data presented in this chapter include specimens and linked data for the first swab per individual (N=2116).

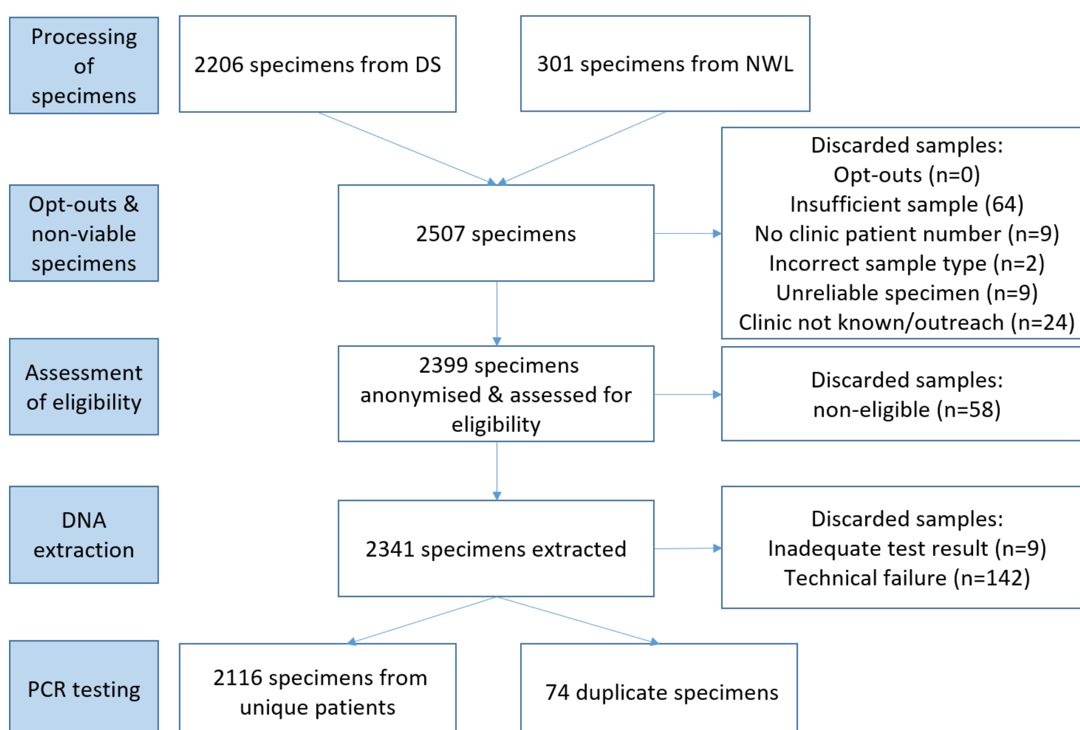


Figure 3.2: Flow chart showing the number of specimens processed, assessed for eligibility and PCR tested

3.4.2 Linkage to socio-demographic, clinical and behavioural data

Of the 2116 specimens from unique study participants, 2107 (99.6%) matched to GUMCAD and nine had no clinical records. Of those who matched, 1991 (94.1%) individuals matched using both the clinic patient number and attendance date, while 125 (5.9%) matched using the clinic patient number only. For the latter, it was assumed that the clinic attendance was not clinically coded in the GUMCAD dataset, however demographic and clinical data were available for other attendances at the clinic from 2008 onwards, and these were used to infer socio-demographic data and prior clinical history for the study population.

Linked clinical and behavioural data (extracted from the clinic database) for the study attendance date were available for 2082 (98.4%) participants (i.e. at least one variable was complete); two participants did not complete the clinical proforma nor did they match to any clinical records in GUMCAD. Of those that did not have any proforma data, nearly half (16/34) had been diagnosed with gonorrhoea in the past 42 days, suggesting that some of these study attendances were for test-of-cure swabs where proforma data were not collected (G Whitlock, personal communication, August 2018). However, it was not possible to confirm this using the data collected in the study.

3.4.3 Description of incomplete data variables

Missing demographic data from the GUMCAD surveillance system ranged from 0.4% (9/2116) for age to 4.6% (97/2116) for region of birth. Data extracted from the clinic patient database had a higher percentage of missing data than GUMCAD (Table 3.3). The percentage of missing behavioural data ranged from 9.8% (208/2116) for receptive anal sex in the past three months to 23.4% (494/2116) for number of new sexual partners in the past three months. Among people who had completed the clinical proforma, 23.7% (493/2082) did not provide any information on symptoms of gastroenteritis or diarrhoeal illness, and this was taken to indicate the absence of symptoms.

Table 3.3: Proportion of missing data for behavioural and clinical variables with more than 5% missing data

Variable	Number (%)
Receptive anal sex (past 3 months)	208 (9.8)
Last condomless sex	234 (11.1)
Receptive oral sex (past 3 months)	242 (11.3)
Current PrEP use (n=1744)	267 (15.3)
Interest in specific high-risk practices	341 (16.1)
Number of sexual partners (past 3 months)	414 (19.6)
Number of new sexual partners (past 3 months)	494 (23.4)

N=2116 unless otherwise specified. 'Interest in specific high-risk practices' refers to data collected via the following question on the clinical proforma: "Are you into any of these? Fisting, injecting, bare backing, chemsex"

Key behavioural data (number of sexual partners, number of new sexual partners, last condomless sex, receptive anal sex, receptive oral sex and 'interest in specific high-risk practices') were complete for 74.7% (1581/2116) of study participants. Despite a high level of completion, there were systematic differences between people with observed and missing data. For instance, people with missing data for all these variables were more likely to have been diagnosed with gonorrhoea in the past 42 days (24.4% vs 3.5%, p<0.001), to have attended 56DS (56.3% vs 16.7%, p<0.001), to be living with HIV (52.6% vs 15.2%, p<0.001), and to be of an ethnic minority group (30.8% vs 21.6%, p=0.015) compared to those with at least one variable completed. The pattern of missing data varied according to the clinic attended and was likely related to the division in service provision across the two clinics. Among 56DS attendees, people with missing data for all variables were more likely to be living with HIV (77.6 vs 19.6%, p<0.001). However, there was no evidence for an association between HIV status and missing data among people attending DSE (20.3% all missing data vs 14.3% with at least one complete variable, p=0.196). This difference may be related to the fact that people living with HIV attend 56DS specifically for HIV care, and a clinical proforma may not be routinely completed during this type of consultation (G Whitlock, personal communication, August 2018). STI swabs may be taken opportunistically at these visits either

without or with only partial completion of the clinical proforma. At both clinics, there was a strong association between a recent (past 42 days) diagnosis of gonorrhoea and missing behavioural data, suggesting that some people may have been attending for test-of-cure swabs where a clinical proforma was not routinely completed (DSE: 30.0% vs 3.7%, p<0.001, 56DS: 15.7% vs 2.5%, p<0.001). When restricted to people who had information for at least one behavioural variable, people with missing data for at least one data item were more likely to have attended 56DS compared to those with complete data. Among people attending DSE, older age and HIV-negative/unknown status were associated with missing data.

These analyses suggested that there were multiple reasons for missing data. Clinical proforma completion was likely related to the reason for attendance as well as the clinic attended. Further discussions with the clinicians at DS also revealed that people taking PrEP may complete a different clinical proforma (G Whitlock, personal communication, September 2019). Missing PrEP status was therefore dependent on whether the individual was taking PrEP and those who were taking PrEP were more likely to have missing data. These analyses suggested that the missing data mechanism in the study may be MNAR.

3.4.4 Characteristics of study participants

Selected characteristics of the 2116 unique study participants are presented in Table 3.4. Overall, 80.8% (1709) of study participants attended DSE and 19.2% (407) attended 56DS. The median age was 32 years (IQR 27 to 39). Most study participants were of white ethnicity (77.8%) and were gay (96.2%). Nearly half of study participants were born in the UK (47.2%). Heterosexual men were retained in the study because rectal swabs are collected based on reported sexual behaviour, rather than sexual identity (see section 3.3.3).

The median number of sexual partners in the past three months was four (IQR 2-9) and the median number of new sexual partners in the same period was three (IQR 1-6). Overall, 17.6% (372/2116) of study participants were living with HIV; 98.4% (366) were diagnosed prior to the study attendance and 1.6% (6) were diagnosed at the study attendance. Among individuals who were HIV-negative or where HIV status was unknown, 75.5% (1316/1739) tested HIV-negative at the study attendance and a further 23.7% (412/1739) had tested HIV-negative

within the past year. Where clinical proforma data were available, 1.7% (36/2082) reported symptoms of gastroenteritis or diarrhoeal illness.

Table 3.4: Socio-demographic, clinical and behavioural characteristics of the study population

Characteristic	Number	Percentage (%)
Age group		
16-19	31	1.5
20-24	241	11.4
25-29	526	25.0
30-34	485	23.0
35-39	337	16.0
40-49	339	16.1
50+	148	7.0
Missing	9	
Ethnic group		
White	1576	77.8
Black	76	3.8
Mixed	131	6.5
Asian	112	5.5
Other	130	6.6
Missing	91	
World region of birth		
UK	953	46.2
Europe	581	28.8
Asia	158	7.8
South America	122	6.0
North America	53	2.6
Central America & the Caribbean	24	1.2
Africa	69	3.4
Australasia	59	3.0
Missing	97	
Sexual orientation		
Gay	2003	96.2
Bisexual	54	2.6
Heterosexual	25	1.2
Missing	34	

Characteristic	Number	Percentage (%)
IMD quintile		
1 (Most deprived)	573	27.5
2	834	40.0
3	374	17.9
4	203	9.7
5 (Least deprived)	101	4.8
Missing	31	
Number of sexual partners (past 3 months)		
0	12	0.7
1	170	10.0
2-4	678	39.8
5-9	440	25.9
10+	402	23.6
Missing	414	
Number of new sexual partners (past 3 months)		
0	172	10.5
1	270	16.7
2-4	588	36.3
5-9	347	21.4
10+	245	15.1
Missing	494	
Receptive anal sex (past 3 months)		
No	92	4.8
Yes	1816	95.2
Missing	208	
Receptive oral sex (past 3 months)		
No	47	2.5
Yes	1827	97.5
Missing	242	
Last condomless sex		
Never	184	9.8
More than 6 weeks ago	437	23.2
Within 6 weeks	986	52.4
Within 72 hours	275	14.6
Missing	234	
Interest in specific high-risk practices*		
No	1074	60.5
Yes	701	39.5
Missing	341	
Bacterial STI diagnosis (at attendance)		
No/unknown	1632	77.1
Yes	484	22.9

Characteristic	Number	Percentage (%)
Bacterial STI diagnosis (past year)		
No/unknown	1251	59.1
Yes	865	40.9
HIV status		
Negative/unknown	1744	82.7
HIV-diagnosed prior to attendance	366	17.3
HIV-diagnosed at attendance	6	0.3
Currently using PrEP (N=1744)		
No	930	63.0
Yes	547	37.0
Missing	267	

N=2116 unless otherwise specified. *'Interest in specific high-risk practices' refers to data collected via the following question: Are you into any of these: Fisting, injecting, bare backing, chemsex.

My literature review (Chapter 2) found that outbreaks of enteric pathogens have occurred in MSM reporting specific sexual practices, including chemsex, and who were living with HIV. Here, I explore the association between 'interest in specific high-risk practices' and HIV status, and I also investigate whether these two variables were associated with other sexual behaviours and a history of bacterial STIs.

Men who reported an 'interest in specific high-risk practices' were more likely to be living with HIV (20.1% vs 12.4%, p<0.001), to report a higher number of sexual partners in the past three months (median 6 [IQR 3-12] vs 4 [IQR 2-6], p<0.001), to report receptive anal sex in the past three months (97.4% vs 93.1%, p<0.001) and to have had a bacterial STI in the last 12 months (47.9% vs 33.8%, p<0.001) compared to those who did not report an 'interest in specific high-risk practices'. These men were also slightly older (median age 33 years [IQR 28-40] for men who reported an 'interest in specific high-risk practices' compared to a median age of 31 years [IQR 25-37] for men who did not report an 'interest in specific high-risk practices', p<0.001).

Compared to men who were HIV-negative or of unknown HIV status, those living with HIV were more likely to report condomless sex in the last six weeks (80.1% vs 64.7%, p<0.001) and slightly more likely to report receptive anal sex (97.6% vs 94.8%, p=0.042), but there was no difference in the proportion that reported receptive oral sex (98.2% vs 97.4%, p=0.382) or in the median number of sexual partners (median 5 [IQR 2-10] vs 4 [IQR 3-9], p=0.863). Men living with HIV were also more likely to have had a bacterial STI in the last 12 months (53.5%

vs 38.2%, p<0.001). Among HIV-negative or unknown status men, those currently taking PrEP were more likely to have a higher number of sexual partners in the past three months (median 6 [IQR 4-15] vs 4 [IQR 2-6], p<0.001), to report an 'interest in specific high-risk practices' (61.2% vs 27.0%, p<0.001), and to report receptive anal (98.4% vs 92.8%, p<0.001) and oral (99.2% vs 96.3%, p=0.001) sex in the past three months compared to those who were not taking PrEP.

3.4.5 Representativeness of study population

The demographic characteristics of the study population broadly mirrored those of all MSM attending DS during the study period, although a slightly higher proportion of study participants were of an ethnic minority group and born outside of the UK (Table 3.5).

Study participants had a higher number of sexual partners in the past three months compared to MSM attending clinics participating in the pilot of new GUMCAD data fields to collect behavioural and clinical data, or MSM in the general population as evidenced by Natsal-3 (see section 3.3.12.3). Among men who reported at least one sexual partner in the past three months, 49.8% (842/1690) of study participants reported five or more sexual partners compared to 19.4% (164/844) of MSM attending GUMCAD pilot clinics, and 14% (17/123) of MSM in the general population in Natsal-3.

Table 3.5: Comparison of socio-demographic characteristics of the study population with all eligible MSM clinic attendees during the study period

		MSM 16+ years attending clinics during the study period*		
Characteristic	Study population n (%) N=2,116	Dean Street n (%) N=5,286	London SHCs n (%) N=12,449	England SHCs n (%) N=22,740
Age group				
16-19	31 (1.5)	51 (1.0)	191 (1.5)	759 (3.4)
20-24	241 (11.4)	608 (11.5)	1,413 (11.4)	3,416 (15.1)
25-29	526 (25.0)	1,291 (24.4)	2,708 (21.8)	4,820 (21.5)
30-34	485 (23.0)	1,193 (22.6)	2,602 (20.9)	4,150 (18.3)
35-39	337 (16.0)	859 (16.3)	1,919 (15.4)	2,995 (13.2)
40-49	339 (16.1)	887 (16.8)	2,276 (18.3)	3,720 (16.4)
50+	148 (7.0)	397 (7.5)	1,335 (10.7)	2,820 (12.4)
Missing	9	0	5	60
Ethnic group				
White	1,576 (77.8)	4,032 (79.0)	8,672 (74.8)	17,169 (80.9)
Black	76 (3.8)	218 (4.3)	776 (6.7)	976 (4.6)
Mixed	131 (6.5)	288 (5.6)	662 (5.7)	970 (4.6)
Asian	112 (5.5)	252 (4.9)	728 (6.3)	1,134 (5.3)
Other	130 (6.4)	313 (6.1)	751 (6.5)	974 (4.6)
Missing	91	183	860	1,517
World region of Birth				
UK	953 (47.2)	2,554 (50.4)	5,809 (51.4)	13,817 (66.3)
Europe	581 (28.8)	1,374 (27.1)	2,886 (25.5)	3,657 (17.6)
Asia	158 (7.8)	362 (7.1)	873 (7.7)	1,246 (6.0)
South America	122 (6.0)	253 (5.0)	615 (5.4)	694 (3.3)
North America	53 (2.6)	158 (3.1)	162 (1.4)	209 (1.0)
Central America & the Caribbean	24 (1.2)	56 (1.1)	273 (2.4)	319 (1.5)
Africa	69 (3.4)	156 (3.1)	386 (3.4)	556 (2.7)
Australasia	59 (3.0)	159 (3.1)	303 (2.7)	342 (1.6)
Missing	97	214	1,142	1,900
Patient residence				
London	1,928 (92.3)	4,782 (90.5)	10,880 (92.6)	10,954 (50.1)
UK elsewhere	162 (7.8)	461 (8.7)	874 (7.4)	10,905 (49.9)
Missing	26	43	695	881

*MSM clinic attendees who had a sample collected for chlamydia and gonorrhoea testing as reported through the GUMCAD surveillance system

3.4.6 Detection of bacterial enteric pathogens

207 out of 2116 men had a BEP detected by PCR giving an estimated overall prevalence in the study population of 9.8% (95% CI: 8.5% to 11.1%). Prevalence was slightly higher among men who attended DSE compared to 56DS (10.2% [95% CI: 8.8% to 11.7%] vs 8.1% [95% CI: 5.5% to 11.2%]) but this was not statistically significant (p=0.206) (Figure 3.3). Prevalence ranged from 0.8% (95% CI: 0.4% to 1.2%) for *Shigella* spp. to 4.9% (95% CI: 4.0% to 5.9%) for EAEC. *Salmonella* spp. were not detected in any specimens. There was no evidence of a statistically significant difference in prevalence by clinic for any of the pathogens detected.

Eleven men (5.3%, 11/207) had more than one BEP detected; one had a positive PCR result for *C. coli* and *C. jejuni*, two were positive for EPEC and *C. jejuni*, two for EAEC and *C. jejuni*, four for EPEC and EAEC, two for STEC and EAEC, and one for *Shigella* spp. and EPEC.

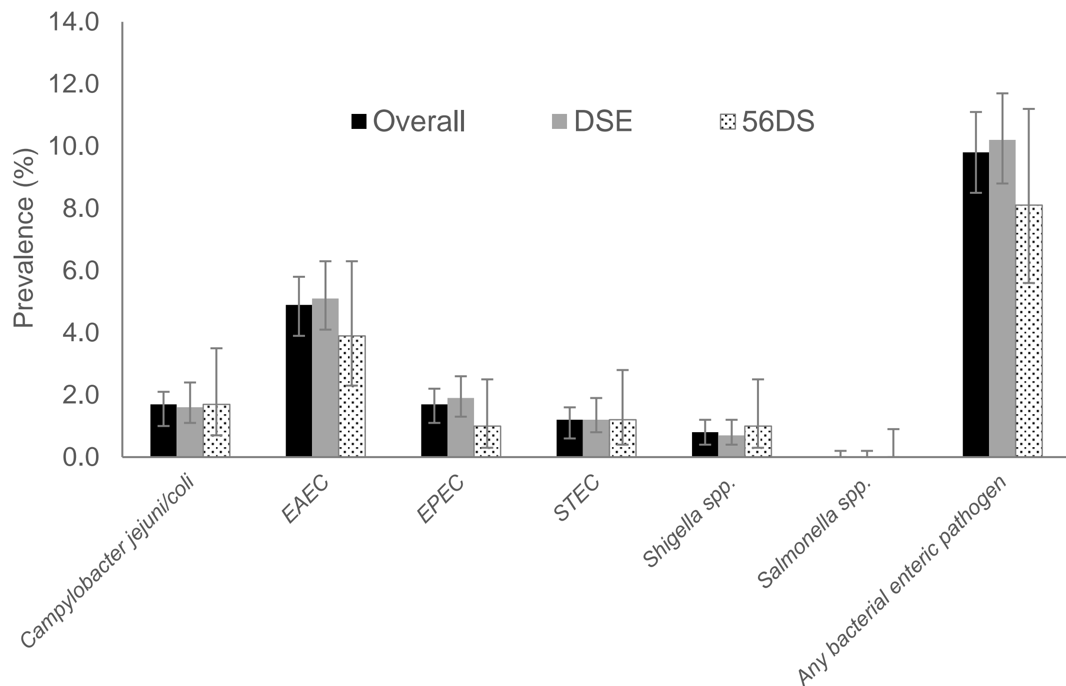


Figure 3.3: Detection of bacterial enteric pathogens by clinic attended

N=2116. Error bars represent 95% Confidence Intervals. EAEC=enteroaggregative *E. coli*, EPEC=enteropathogenic *E. coli*, STEC=Shiga toxin-producing *E. coli*

3.4.7 Risk factors for bacterial enteric pathogen detection

There was no evidence of an association between the detection of a BEP and socio-demographic factors (ethnic group, region of birth, IMD quintile and sexual orientation), except that men of mixed ethnicity had a lower prevalence than men of white ethnicity (adjusted prevalence ratio [aPR]: 0.37 [95% CI: 0.16 to 0.88]) although the sample size for men of mixed ethnicity was small (n=5/131) (Table 3.6). In univariable analysis, the linear test for trend for age group suggested weak evidence for an increase in BEP prevalence by increasing age group (p=0.068), but this was no longer significant after adjustment for clinic and HIV status (p=0.112). Where reported, symptoms of gastroenteritis were not significantly associated with the detection of a BEP, although again the sample size was small (n=5/36).

In univariable analyses, the detection of any BEP was positively associated with markers of higher-risk sexual behaviour including increasing numbers of new (linear test for trend p<0.001) or total (linear test for trend p<0.001) sexual partners in the previous three months, a diagnosis of a bacterial STI at the same attendance date (PR: 1.44 [95% CI: 1.09 to 1.91], p=0.010) or in the previous 12 months (PR: 1.46 [95% CI: 1.13 to 1.89], p=0.004) and reporting an 'interest in specific high-risk practices' (PR: 1.34 [95% CI: 1.02 to 1.77], p=0.036) (Table 3.6). There was also weak evidence to suggest that recent condomless sex was associated with the detection of a BEP (p=0.062), primarily due to higher prevalence among men who reported condomless sex within the last 72 hours. Compared to men who reported that they had never engaged in condomless sex or who had done so more than six weeks ago, the PR was 1.15 (95% CI: 0.85 to 1.58) for men who reported condomless sex within the last six weeks and 1.59 (95% CI: 1.08 to 2.35) for those who reported condomless sex within the last 72 hours.

After adjusting for age, clinic and HIV status, there was little or no change in the PR for numbers of new or total sexual partners in the previous three months, and a diagnosis of a bacterial STI at the same attendance or within the previous 12 months, and these variables remained significantly associated with the detection of a BEP (Table 3.6). Prevalence of a BEP was more than two-fold higher among men who reported 10 or more new sexual partners compared to those who reported none or one new sexual partner (aPR: 2.42 [95% CI: 1.53 to

3.83], p<0.001). On the other hand, there was only weak evidence for an association between detection of a BEP and ‘interest in specific high-risk practices’ (aPR: 1.28 [95% CI: 0.96-1.71], p=0.087) and recent condomless sex (p=0.090) after adjusting for age, clinic and HIV status. Compared to men who reported that they had never engaged in condomless sex or who had done so more than six weeks ago, the aPR was 1.12 (95% CI: 0.81 to 1.54) for men who reported condomless sex within the last six weeks and 1.53 (95% CI: 1.03 to 2.27) for those who reported condomless sex within the last 72 hours.

HIV status was not associated with the detection of a BEP when included as a dichotomous variable (PR: 1.27 [95% CI: 0.92 to 1.73] for men living with HIV compared to men who were of HIV-negative/unknown HIV status, p=0.141). However, I explored this further by creating a categorical variable which included stratification by current PrEP use (‘HIV risk group’ in Table 3.6). I found that HIV-negative men taking PrEP and men living with HIV were more likely to have a BEP detected compared to HIV-negative men who were not taking PrEP (PR: 2.10 [95% CI: 1.52 to 2.90] for HIV-negative men taking PrEP and PR: 1.83 [95% CI: 1.27 to 2.65] for men living with HIV, p<0.001). HIV risk group remained strongly associated with BEP detection after adjustment for age and clinic. Compared to HIV-negative men who were not taking PrEP, the aPR was 2.07 (95% CI: 1.49 to 2.88) for HIV-negative men taking PrEP and 1.88 (95% CI: 1.27 to 2.78) for men living with HIV (p<0.001).

3.4.7.1 Sensitivity analyses

Sensitivity analyses using simple and multiple imputation methods were conducted for the risk factor analyses described above to assess the potential bias of missing data on the results (Appendices 3.8 to 3.12). These sensitivity analyses supported the findings presented in the primary results and strengthen their validity:

- Using the missing indicator method, the strength of association between BEP detection and each behavioural variable was similar to the primary analyses (Appendix 3.8).
- Replacing missing behavioural data with the lowest observed value attenuated the measures of association for most variables, but the same factors remained associated with a prevalent BEP (Appendix 3.9). On the other hand, in this model, the strength of

association was accentuated for ‘interest in specific high-risk practices’ and last condomless sex.

- Replacing missing values with the highest observed value for each behavioural variable somewhat attenuated the strength of association in the highest category (Appendix 3.10). Except for last condomless sex and ‘interest in specific high-risk practices’ where there was no evidence for an association, the same factors remained associated with the detection of a BEP.
- Higher partner number was strongly associated with the detection of a BEP after replacing missing values with the median value (Appendix 3.11). The associations remained after adjusting for age, clinic and HIV status.
- The estimates generated using multiple imputation were similar to the primary analyses and the same variables remained strongly associated with BEP detection. There was some attenuation in the measures of association for number of new sexual partners, whereas the measures of association for total number of sexual partners were slightly accentuated. For last condomless sex and ‘interest in specific high-risk practices’, there was no evidence for an association with BEP detection (Appendix 3.12).

Table 3.6: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic (N=2116)						
DSE	174/1709	10.2	1.00	0.210	1.00	0.138
56DS	33/407	8.1	0.80 (0.56-1.14)		0.76 (0.53-1.09)	
Age group (N=2107)						
16-24	19/272	7.0	1.00	0.173	1.00	0.268
25-34	98/1011	9.7	1.39 (0.86-2.23)	0.068[a]	1.35 (0.84-2.17)	0.112[a]
35+	90/824	10.9	1.56 (0.97-2.52)		1.49 (0.92-2.43)	
Ethnic group (N=2025)						
White	169/1576	10.7	1.00	0.161	1.00	0.193
Black	8/76	10.5	0.98 (0.50-1.92)		0.97 (0.50-1.90)	
Mixed	5/131	3.8	0.36 (0.15-0.85)		0.37 (0.16-0.88)	
Asian	10/112	8.9	0.83 (0.45-1.53)		0.87 (0.47-1.60)	
Other	10/130	7.7	0.72 (0.39-1.32)		0.72 (0.39-1.33)	
Region of birth (N=2019)						
UK	90/953	9.4	1.00	0.546	1.00	0.552
Europe	65/581	11.2	1.18 (0.87-1.60)		1.18 (0.87-1.60)	
Rest of world	49/485	10.1	1.07 (0.77-1.49)		1.06 (0.76-1.47)	
IMD quintile (N=2085)						
1-2 (Most deprived)	144/1407	10.2	1.00	0.611	1.00	0.589
3	33/374	8.8	0.86 (0.60-1.24)		0.86 (0.60-1.23)	
4-5 (Least deprived)	27/304	8.9	0.87 (0.59-1.28)		0.86 (0.58-1.28)	
Sexual orientation (N=2082)						
Gay	200/2003	10.0	1.00	0.298	1.00	0.344
Bisexual/heterosexual	5/79	6.3	0.63 (0.27-1.50)		0.66 (0.28-1.56)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
HIV status (N=2116)						
HIV-negative/unknown	163/1744	9.4	1.00	0.141	1.00	0.162
Living with HIV	44/372	11.8	1.27 (0.92-1.73)		1.26 (0.91-1.76)	
HIV risk group (N=1849)						
HIV-negative/unknown HIV status, not on PrEP	60/930	6.5	1.00	<0.001	1.00	<0.001
HIV-negative, on PrEP	74/547	13.5	2.10 (1.52-2.90)		2.07 (1.49-2.88)	
Living with HIV	44/372	11.8	1.83 (1.27-2.65)		1.88 (1.27-2.78)	
Bacterial STI diagnosed at attendance (N=2116)						
No/unknown	145/1632	8.9	1.00	0.010	1.00	0.010
Yes	62/484	12.8	1.44 (1.09-1.91)		1.45 (1.09-1.91)	
Bacterial STI diagnosed in past year (N=2116)						
No/unknown	103/1251	8.2	1.00	0.004	1.00	0.010
Yes	104/865	12.0	1.46 (1.13-1.89)		1.41 (1.08-1.84)	
Interest in specific high-risk practices (N=1775)						
No	97/1074	9.0	1.00	0.036	1.00	0.087
Yes	85/701	12.1	1.34 (1.02-1.77)		1.28 (0.96-1.71)	
Number of sexual partners in last 3 months (N=1702)						
0-1	14/182	7.7	1.00	0.005	1.00	0.009
2-4	52/678	7.7	0.99 (0.57-1.76)	<0.001[a]	0.97 (0.55-1.72)	0.013[a]
5-9	47/440	10.7	1.39 (0.78-2.46)		1.34 (0.76-2.38)	
10+	57/402	14.2	1.84 (1.06-3.22)		1.75 (1.00-3.07)	
Number of new sexual partners in last 3 months (N=1622)						
0-1	28/442	6.3	1.00	<0.001	1.00	<0.001
2-4	54/588	9.2	1.45 (0.93-2.25)	<0.001[a]	1.47 (0.95-2.26)	<0.001[a]
5-9	45/347	1.0	2.05 (1.30-3.21)		2.03 (1.30-3.17)	
10+	38/245	15.5	2.45 (1.54-3.89)		2.42 (1.53-3.83)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Receptive anal sex in last 3 months (N=1908)						
No	7/92	7.6	1.00	0.475	1.00	0.555
Yes	180/1816	9.9	1.30 (0.63-2.69)		1.24 (0.61-2.54)	
Receptive oral sex in last 3 months (N=1874)						
No	6/47	12.8	1.00	0.486	1.00	0.413
Yes	178/1827	9.7	0.76 (0.36-1.63)		0.73 (0.35-1.55)	
Last condomless sex (N=1882)						
Never/more than 6 weeks ago	54/621	8.7	1.00	0.062	1.00	0.090
Within 6 weeks	99/986	10.0	1.15 (0.85-1.58)		1.12 (0.81-1.54)	
Within 72 hours	38/275	13.8	1.59 (1.08-2.35)		1.53 (1.03-2.27)	
Gastrointestinal symptoms (N=2082)						
No/unknown	201/2046	9.8	1.00	0.410	1.00	0.195
Yes	5/36	13.9	1.41 (0.62-3.22)		1.76 (0.75-4.13)	

Total numbers vary for each question due to missing data. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Overall p-values by Wald test or linear test for trend ([a]). Adjusted Models: Each factor adjusted in a separate model for age (continuous variable), clinic and HIV status. aPRs and p-value presented for age group for ease of interpretation. 'Interest in specific high-risk practices' refers to data collected via the following question: Are you into any of these: Fisting, injecting, bare backing, chemsex.

3.4.7.2 Risk factors for BEP detection by HIV status

There was evidence that HIV status modified the effect of 'interest in specific high-risk practices' (p=0·001) and the number of new sexual partners (p=0.035) on BEP detection. As a result, the following section presents the risk factor analyses separately for HIV-negative/unknown HIV status MSM and for those who were living with HIV.

3.4.7.2.1 Risk factors for BEP detection in HIV-negative/unknown status MSM

Amongst MSM who were HIV-negative or of unknown HIV status, there was no evidence for an association between socio-demographic factors and the detection of a BEP, except that men of mixed ethnicity were less likely to have a BEP detected compared to those of white ethnicity (aPR: 0.28 [95% CI: 0.09 to 0.86]). The detection of a BEP was associated with a suite of higher risk sexual behaviours in both univariable and multivariable analyses (Table 3.7). Moreover, reporting an 'interest in specific high-risk practices' was strongly associated with the detection of a BEP, after adjusting for age and clinic (aPR: 1.65 [95% CI: 1.21 to 2.26], p=0.002). There was evidence to suggest that recent condomless sex was associated with the detection of a BEP (p=0.050), although this was of borderline significance. Compared to men who reported that they had never engaged in condomless sex or who had done so more than six weeks ago, the aPR was 1.04 (95% CI: 0.73 to 1.47) for men who reported condomless sex within the last six weeks and 1.61 (95% CI: 1.06 to 2.45) for those who reported condomless sex within the last 72 hours.

3.4.7.2.2 Sensitivity analyses in HIV-negative/unknown status MSM

I conducted sensitivity analyses to assess the impact of missing data on the findings. The results from these sensitivity analyses were similar to those from the main analysis (Appendices 3.13 to 3.16):

- Using the missing indicator method, the findings supported those presented in the main analyses and the direction of the association was similar (Appendix 3.13).
- Replacing missing behavioural data with the lowest observed value resulted in some or no attenuation to the PR, and the same factors remained associated with BEP detection. Compared to the main analysis, there was stronger evidence to suggest that recent

condomless sex was associated with BEP detection and the PRs were slightly accentuated (Appendix 3.14).

- Replacing missing behavioural data with the highest observed value attenuated the prevalence ratio for the highest category. Compared to the main analysis, the same variables remained associated with BEP detection, except for last condomless sex where there was no evidence for an association with BEP detection (Appendix 3.15).
- Higher partner number was strongly associated with the detection of a BEP after replacing missing values with the median value (Appendix 3.16).

Table 3.7: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in HIV-negative/unknown status MSM

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic (N=1744)						
DSE	140/1461	9.6	1.00	0.445	1.00	0.442
56DS	23/283	8.1	0.85 (0.56-1.29)		0.85 (0.56-1.29)	
Age group (N=1736)						
16-24	17/262	6.5	1.00	0.129	1.00	0.131
25-34	80/868	9.2	1.42 (0.86-2.35)	0.043[a]	1.41 (0.85-2.35)	0.044[a]
35+	66/606	10.9	1.68 (1.00-2.80)		1.67 (1.00-2.80)	
Ethnic group (N=1667)						
White	136/1304	10.4	1.00	0.149	1.00	0.180
Black	5/56	8.9	0.86 (0.37-2.01)		0.86 (0.37-2.02)	
Mixed	3/108	2.8	0.27 (0.09-0.82)		0.28 (0.09-0.86)	
Asian	8/96	8.3	0.80 (0.40-1.58)		0.83 (0.42-1.65)	
Other	7/103	6.8	0.65 (0.31-1.36)		0.66 (0.32-1.37)	
Region of birth (N=1664)						
UK	71/802	8.9	1.00	0.549	1.00	0.537
Europe	51/477	10.7	1.21 (0.86-1.70)		1.21 (0.86-1.71)	
Rest of world	38/385	9.9	1.11 (0.77-1.62)		1.11 (0.76-1.61)	
IMD quintile (N=1714)						
1-2 (Most deprived)	108/1149	9.4	1.00	0.990	1.00	0.969
3	29/313	9.3	0.99 (0.67-1.46)		0.98 (0.66-1.44)	
4-5 (Least deprived)	23/252	9.1	0.97 (0.63-1.49)		0.95 (0.61-1.47)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Sexual orientation (N=1717)						
Gay	157/1641	9.6	1.00	0.394	1.00	0.395
Bisexual/heterosexual	5/76	6.6	0.69 (0.29-1.63)		0.69 (0.29-1.63)	
PrEP use (N=1477)						
No	60/930	6.5	1.00	<0.001	1.00	<0.001
Yes	74/547	13.5	2.10 (1.52-2.90)		2.06 (1.49-2.87)	
Bacterial STI diagnosed at attendance (N=1744)						
No/unknown	116/1357	8.6	1.00	0.031	1.00	0.027
Yes	47/387	12.1	1.42 (1.03-1.96)		1.43 (1.04-1.97)	
Bacterial STI diagnosed in previous year (N=1744)						
No/unknown	85/1078	7.9	1.00	0.008	1.00	0.009
Yes	78/666	11.7	1.49 (1.11-1.99)		1.47 (1.10-1.97)	
Interest in specific high-risk practices (N=1501)						
No	72/941	7.7	1.00	0.001	1.00	0.002
Yes	72/560	12.9	1.68 (1.23-2.29)		1.65 (1.21-2.26)	
Number of sexual partners in last 3 months (N=1416)						
0-1	12/148	8.1	1.00	0.003	1.00	0.005
2-4	35/570	6.1	0.76 (0.40-1.42)	<0.001[a]	0.72 (0.39-1.35)	0.011[a]
5-9	41/370	11.1	1.37 (0.74-2.53)		1.28 (0.70-2.37)	
10+	44/328	13.4	1.65 (0.90-3.04)		1.54 (0.83-2.83)	
Number of new sexual partners in last 3 months (N=1351)						
0-1	18/355	5.1	1.00	<0.001	1.00	<0.001
2-4	41/505	8.1	1.60 (0.94-2.74)	<0.001[a]	1.59 (0.94-2.70)	<0.001[a]
5-9	38/286	13.3	2.62 (1.53-4.49)		2.59 (1.52-4.41)	
10+	32/205	15.6	3.08 (1.77-5.34)		2.99 (1.73-5.17)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Receptive anal sex in last 3 months (N=1622)						
No	5/85	5.9	1.00	0.299	1.00	0.304
Yes	143/1537	9.3	1.58 (0.67-3.76)		1.57 (0.67-3.67)	
Receptive oral sex in last 3 months (N=1590)						
No	4/42	9.5	1.00	0.926	1.00	0.884
Yes	141/1548	9.1	0.96 (0.37-2.46)		0.93 (0.36-2.40)	
Last condomless sex (N=1605)						
Never or more than 6 weeks ago	49/566	8.7	1.00	0.045	1.00	0.050
Within 6 weeks	73/811	9.0	1.04 (0.74-1.47)		1.04 (0.73-1.47)	
Within 72 hours	32/228	14.0	1.62 (1.07-2.46)		1.61 (1.06-2.45)	
Gastrointestinal symptoms (N=1715)						
No/unknown	159/1686	9.4	1.00	0.867	1.00	0.701
Yes	3/29	10.3	1.10 (0.37-3.24)		1.25 (0.41-3.84)	

Total numbers vary for each question due to missing data. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Overall p-values by Wald test or linear test for trend ([a]). Adjusted Models: Each factor adjusted in separate model for age (continuous variable) and clinic. aPRs and p-value presented for age group for ease of interpretation. 'Interest in specific high-risk practices' refers to data collected via the following question: Are you into any of these: Fisting, injecting, bare backing, chemsex.

3.4.7.2.3 Risk factors for BEP detection in MSM living with HIV

Amongst MSM living with HIV, most variables of sexual behaviour were not associated with the detection of a BEP in univariable analyses, or in the multivariable strategy adjusting for age and clinic (Table 3.8).

After adjusting for age and clinic, men who reported symptoms of gastroenteritis were more likely to have a BEP detected (aPR: 4.06 [95% CI 1.05 to 15.7], p=0.042), but the confidence interval was wide and the sample size was small (n=2/7). An unexpected finding in the group of MSM living with HIV was that men who reported an 'interest in specific high-risk practices' were significantly less likely to have a BEP detected compared to those who did not report this (aPR: 0.49 [95% CI: 0.26 to 0.93], p=0.029). In addition, men who reported receptive oral sex in the last three months were less likely to have a BEP detected, although this was of borderline significance and the sample size for those who did not report receptive oral sex was small (n=2/5) (aPR: 0.33 [95% CI: 0.11 to 1.02], p=0.054).

3.4.7.2.4 Sensitivity analyses in MSM living with HIV

As for HIV-negative/unknown status MSM, I performed sensitivity analyses to assess the impact of missing data on the primary results. The results from the sensitivity analyses were mixed and for some variables, there were conflicting findings (Appendices 3.17 to 3.20):

- Using the missing indicator method, men who reported an 'interest in specific high-risk practices' were less likely to have a BEP detected, as reported in the primary analyses. However, there was weaker evidence to suggest that men who engaged in receptive oral sex in the past three months were less likely to have a BEP detected, after adjusting for age and clinic (aPR: 0.33 [95% CI: 0.10 to 1.05], p=0.069). There was no evidence for any difference in the prevalence of a BEP according to reported symptoms (Appendix 3.17).
- After replacing missing values with the lowest observed value, there was little evidence to suggest that men who reported an 'interest in specific high-risk practices' were less likely to have a BEP detected (aPR: 0.57 [95% CI: 0.29 to 1.11], p=0.098). MSM who had symptoms of gastroenteritis were more likely to have a BEP detected, after adjusting for age and clinic (aPR: 4.13 [95% CI: 1.07-16.0], p=0.040) (Appendix 3.18).

- After replacing missing values with the highest observed value, there was strong evidence to suggest that men who reported an ‘interest in specific high-risk practices’ were less likely to have a BEP detected (aPR: 0.45 [95% CI: 0.26-0.81], p=0.007) (Appendix 3.19). Men who reported receptive oral sex in the past three months were less likely to have a BEP detected, although this was of borderline significance after adjusting for age and clinic (aPR: 0.30 [95% CI: 0.09 to 1.03], p=0.056). There was no evidence for any difference in the prevalence of a BEP according to reported symptoms of gastroenteritis.
- There was no evidence for an association between partner number and the detection of a BEP after replacing the missing values with the median number of sexual partners (Appendix 3.20).

Table 3.8: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in MSM living with HIV

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic (N=372)						
DSE	34/248	13.7	1.00	0.122	1.00	0.145
56DS	10/124	8.1	0.59 (0.30-1.15)		0.60 (0.30-1.19)	
Age group (N=371)						
16-34	20/153	13.1	1.00	0.546	1.00	0.609
35+	24/218	11.0	0.84 (0.48-1.47)		0.86 (0.49-1.52)	
Ethnic group (N=358)						
White	33/272	12.1	1.00	0.901	1.00	0.906
Ethnic minority	10/86	11.6	0.96 (0.49-1.86)		0.96 (0.50-1.85)	
Region of birth (N=355)						
UK	19/151	12.6	1.00	0.865	1.00	0.879
Europe	14/104	13.5	1.07 (0.56-2.04)		1.01 (0.52-1.94)	
Rest of world	11/100	11.0	0.87 (0.43-1.76)		0.85 (0.43-1.70)	
IMD quintile (N=371)						
1-2 (Most deprived)	36/258	14.0	1.00	0.192	1.00	0.195
3	4/61	6.6	0.47 (0.17-1.27)		0.47 (0.17-1.25)	
4-5 (Least deprived)	4/52	7.7	0.55 (0.20-1.48)		0.58 (0.22-1.55)	
Sexual orientation (N=365)						
Gay	43/362	11.9	N/A			
Bisexual/heterosexual	0/3	0				
Bacterial STI diagnosed at attendance (N=372)						
No/unknown	29/275	10.6	1.00	0.195	1.00	0.209
Yes	15/97	15.5	1.47 (0.82-2.62)		1.45 (0.81-2.57)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Bacterial STI diagnosed in last year (N=372)						
No/unknown	18/173	10.4	1.00	0.431	1.00	0.683
Yes	26/199	13.1	1.26 (0.71-2.21)		1.13 (0.63-2.05)	
Interest in specific high-risk practices (N=274)						
No	25/133	18.8	1.00	0.026	1.00	0.029
Yes	13/141	9.2	0.49 (0.26-0.92)		0.49 (0.26-0.93)	
Number of sexual partners in last 3 months (N=286)						
0-4	19/142	13.4	1.00	0.963	1.00	0.973
5+	19/144	13.2	0.99 (0.54-1.78)		0.99 (0.54-1.81)	
Number of new sexual partners in last 3 months (N=271)						
0-4	23/170	13.5	1.00	0.878	1.00	0.892
5+	13/101	12.9	0.95 (0.50-1.80)		0.96 (0.50-1.81)	
Receptive anal sex in last 3 months (N=286)						
No	2/7	28.6	1.00	0.214	1.00	0.184
Yes	37/279	13.3	0.46 (0.14-1.56)		0.43 (0.12-1.50)	
Receptive oral sex in last 3 months (N=284)						
No	2/5	40.0	1.00	0.053	1.00	0.054
Yes	37/279	13.3	0.33 (0.11-1.01)		0.33 (0.11-1.02)	
Last condomless sex (N=268)						
Never/more than 6 weeks ago	5/55	9.1	1.00	0.314	1.00	0.321
Within 6 weeks	32/222	14.4	1.59 (0.65-3.89)		1.58 (0.64-3.93)	
Gastrointestinal symptoms (N=367)						
No/unknown	42/360	11.7	1.00	0.146	1.00	0.042
Yes	2/7	28.6	2.45 (0.73-8.19)		4.06 (1.05-15.7)	

Total numbers vary for each question due to missing data. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Overall p-values by Wald test. Adjusted Models: Each factor adjusted in separate model for age (continuous variable) and clinic. aPRs and p-value presented for age group for ease of interpretation. 'Interest in specific high-risk practices' refers to data collected via the following question: Are you into any of these: Fisting, injecting, bare backing, chemsex.

3.4.8 Detection of *mphA* and its association with a previous bacterial STI diagnosis

All 207 specimens that had a BEP detected underwent real-time PCR detection for the presence of *mphA*, a genotypic marker of azithromycin resistance, alongside a control group of a randomly selected sub-set of 100 specimens that did not have a BEP detected. Among these 307 specimens, one specimen with a BEP detected had insufficient DNA volume (<2.5μl) to enable *mphA* detection, and a further specimen from the control group generated an inadequate PCR result.

I detected *mphA* in 32.5% (99/305) of specimens overall; the detection of *mphA* was more likely in specimens that had a BEP detected compared to the control group (41.3% [85/206] vs 14.1% [14/99], p<0.001). Overall, *mphA* was more likely to be detected in men with a bacterial STI diagnosed in the past year; 41.3% (59/143) of specimens from men with a bacterial STI in the past year had *mphA* detected compared to 24.7% (40/162) of specimens from men without a bacterial STI (p=0.002) (Figure 3.4). Among the sub-group of MSM with a BEP detected, *mphA* was detected in 51.5% (53/103) of specimens from those who had a bacterial STI diagnosis in the past year, but in only 31.1% (32/103) of specimens from those who did not (p=0.003). Among the control group without a BEP, *mphA* was detected in 15.0% (6/40) of specimens from men with a bacterial STI in the past year and in 13.6% (8/59) of specimens from those who did not (p=0.840) (Figure 3.4).

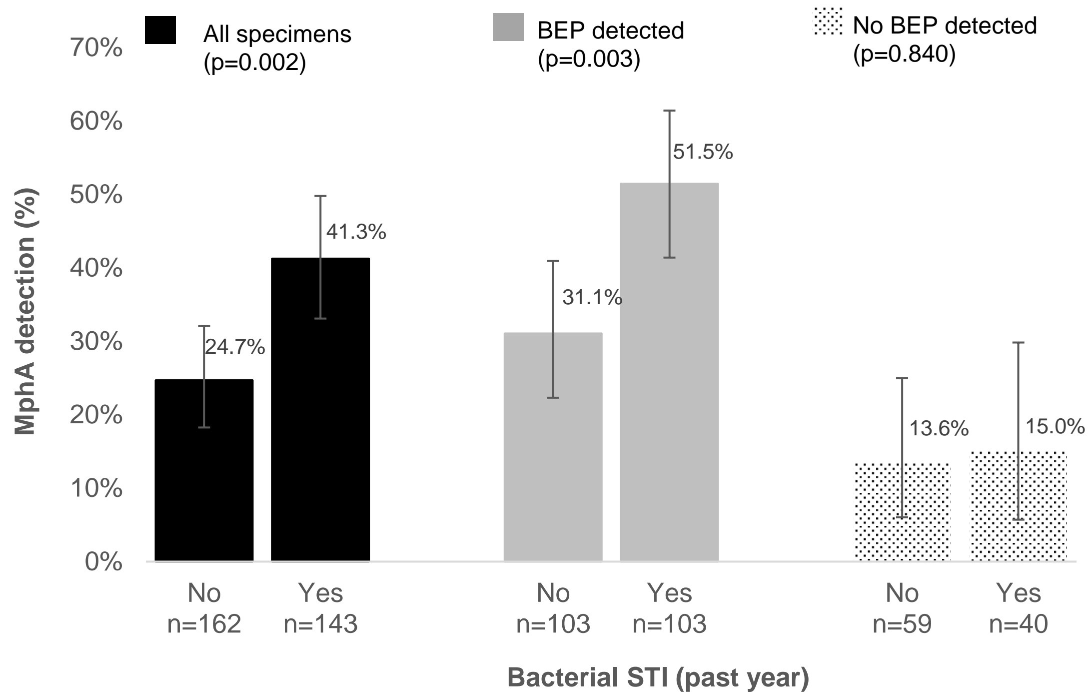


Figure 3.4: Detection of *mphA* by bacterial STI history in the past year stratified by bacterial enteric pathogen detection

p-values generated using Pearson's Chi-squared test

3.5 Discussion

3.5.1 Summary of key findings

In this study, approximately 10% of MSM attending DS had a BEP detected and most of these men reported no symptoms of gastroenteritis or diarrhoeal illness. The detection of a BEP was associated with a suite of higher-risk sexual behaviours, supporting the idea that these pathogens are being transmitted directly through sexual contact. I also found that the detection of *mphA*, a genotypic marker for azithromycin resistance, was more common in specimens where a BEP was detected, but was also present in 14.1% of those without a BEP. Among individuals with a BEP detected, *mphA* was strongly associated with a bacterial STI diagnosis in the past year, which might reflect previous antimicrobial exposure acting as a selective pressure.

3.5.2 Strengths and limitations

For the first time, my study links data on BEP detection to behavioural, clinical and socio-demographic data for a large, predominantly asymptomatic sample of MSM in England, and explores the prevalence of a genetic marker for azithromycin resistance in this population. The main strength of my study was the large and unselected sample of MSM attending the UK's largest SHC, regardless of symptoms. In 2016, DS accounted for 20% of all attendances by MSM attending SHCs in England. I used residual specimens, an opt-out approach and anonymous testing to reduce the potential for selection bias that may have been introduced through participant refusal. Both approaches maximised the proportion of MSM attending the clinic who were included in the study. Overall, there are very few studies that have explored the prevalence of BEPs in MSM and none that have used such a large sample. My study provides the most robust estimates of BEP prevalence among MSM in England to date and are a considerable addition to the literature. To my knowledge, this is also the first study to explore the prevalence of *mphA* and its relationship with BEPs and bacterial STI diagnoses.

The study population was broadly representative of MSM who attended DS during the study period. However, people who attend DS are, on average, at higher risk of STIs and HIV than MSM attending other SHCs or MSM in the general population, as demonstrated by the external data comparisons. In addition, DS reports one fifth of all new HIV and one third of gonorrhoea diagnoses made across England. It is likely that my study sample includes MSM who are at the highest risk of acquiring STIs and enteric pathogens. As a result, my findings may not be generalisable to all MSM.

Although some helpful insight is available from the IID studies to enable comparison of my results to prevalence estimates in the general population, differences in the characteristics of study population, type of specimen and testing methodology need to be taken into consideration (see section 3.5.3.1 below). To aid interpretation, heterosexual men attending SHCs would have provided a useful comparison group to contextualise my findings. However, rectal swabs are not routinely collected from heterosexual men attending SHCs.

This study used rectal swabs, rather than stool specimens, which are the recommended specimen for clinical diagnosis of BEPs. Although I had good feasibility data showing that rectal swabs could be used to detect BEPs, I did not undertake like-for-like comparisons in the same individual and it may be that the use of rectal swabs led to an underestimation of BEP prevalence. Although not directly comparable, Kotar *et al.* (2019) evaluated the diagnostic performance of 304 rectal swabs compared to paired stool samples for the detection of enteric pathogens in adults with diarrhoea. Compared to BEP detection from stool samples, the authors found that the sensitivity of rectal swabs using PCR detection was 86.5% (95% CI: 79.5% to 91.8%).[250] However, this study used a different type of rectal swab, DNA extraction protocol and PCR-based method to those used in my study, and these factors may influence pathogen detection. Another factor to consider is that PCR detection of enteric pathogens is more sensitive than conventional microbiological methods (e.g. bacteriological culture), and to avoid over-estimating prevalence, I selected cut-off values for defining pathogen detection used by the reference laboratory for defining a clinically meaningful result. It is possible that individuals with low pathogen load were reported as negative, which would have underestimated asymptomatic carriage.

Due to the opt-out design, the study dataset was restricted to data items routinely collected at the clinic. I was not able to collect specific information about gastrointestinal symptoms, recent antimicrobial exposure, recent travel history, occupational exposure, food consumption, knowledge of BEP transmission, or more specific information about sexual practices facilitating faecal oral transmission, such as oral-anal contact (rimming) or recent chemsex drug-use, all of which could have helped interpretation. The collection of these data items would have required an opt-in approach and use of a self-administered bespoke study questionnaire.

The routinely collected data used in my study may be subject to information bias resulting from systematic differences in the way that the independent variables (or exposures) were obtained. Information bias can occur due to inaccurate recall or reporting, or missing responses. For example, the data suggested that most men attending the clinic did not have gastrointestinal symptoms, however, I assumed that all men who attended DSE and who had completed the clinical proforma were asymptomatic unless they specifically recorded gastrointestinal

symptoms in the free text field during check-in. This may have led to underestimation of symptoms. All men attending 56DS were asked about symptoms, but it is possible that gastrointestinal symptoms were missed or not recorded accurately, particularly if other STI-related symptoms prevailed. I also used GUMCAD data to infer clinical history in GUMCAD (i.e. STI diagnoses and HIV status). A major limitation of GUMCAD data is the inability to follow patients between SHCs. Consequently, I could have underestimated the number of study participants who had a bacterial STI diagnosis in the past year or who were living with HIV if these diagnoses were reported by a different SHC. Data from a small number of clinics participating in a GUMCAD pilot collecting behavioural and clinical data items estimated that 9% of MSM had attended another SHC in the past year.[247] The impact of movement between SHCs on defining HIV status in my study was, however, likely minimal given that I used data from both GUMCAD and the clinic database at DS. Data on antimicrobial use were also not available for this study and so I used a diagnosis of a bacterial STI in the past 12 months as a proxy measure of previous antimicrobial exposure. Finally, the behavioural data in my study may also be subject to social desirability bias if participants provided inaccurate responses that they considered to be more socially acceptable when providing information to the SHC staff. The use of an anonymised self-administered questionnaire that can be completed in a private space can minimise social desirability bias.[251] Although the use of a questionnaire was explored during the study design phase, the opt-out design with the use of routine data was considered to be the most pragmatic approach (see section 3.3.1).

For some variables, there was a high proportion of missing data, which could have resulted in biased estimates in the risk factor analysis. To better understand the biases, I performed sensitivity analyses to assess the impact of the missing data. While these methods are themselves subject to bias,[245] the results of the sensitivity analyses were generally concordant with the main results. It should be noted that replacing missing values with either the lowest or highest observed value for a variable meant that the data were analysed as if these individuals all belonged to the lowest or highest group, which might not be a realistic scenario. In addition, implementation of multiple imputation in standard statistical packages assumes that the data are MAR.[241] I included a range of variables in the imputation models to increase the plausibility of the MAR assumption. However, as discussed in section 3.4.3, it is possible

that the data in my study were MNAR, which could result in biased estimates when using this imputation method. For variables that were of borderline significance in the main results, sensitivity analyses generated conflicting results. In general, the sensitivity analyses attenuated the measures of association, suggesting that the main analyses may have overestimated the measures of effect. However, the direction of association was similar in the sensitivity analyses for the study population overall and for HIV-negative/unknown status MSM, suggesting that the findings were valid. In addition, the association between BEP and markers of higher-risk sexual behaviour are biologically plausible, since it seems reasonable to assume that this variable would include men who may be more likely to engage in practices that increase the likelihood of oral-anal contact.

The inconsistent findings in the sub-analyses involving MSM living with HIV should be interpreted carefully due to the small sample size. In an underpowered analysis, the proportion of significant findings that appear just by chance will be higher, resulting in type I error (i.e. false positive finding).[252]

3.5.3 Interpretation of results

3.5.3.1 Comparison to prevalence estimates in the general population

My study provides evidence that a range of BEPs are present in the gut of MSM attending DS, including pathogens associated with recent outbreaks in the UK and elsewhere. EAEC was the most common pathogen detected. The prevalence of *Shigella* spp. was low at 0.8% (95% CI: 0.4% to 1.2%) and this was not unexpected given the large drop in the number of cases in adult males with no foreign travel history reported to national surveillance at the time of conducting this study (December 2017 to February 2018). The detection of other BEPs such as *Campylobacter* spp. is important as the large burden of food-borne cases[36,253] at a national level may mask transmission occurring through sex between men.

It is important to acknowledge that these pathogens might also be detected in the gut of many adults unrelated to sexual behaviour. To provide context for my findings, I have compared the prevalence of BEPs among MSM in my study sample to prevalence estimates for BEPs from adults recruited to the first and second studies of 'Infectious Intestinal Disease (IID) in the

Community' (IID1 and IID2). These studies estimated the burden and aetiology of enteric infections at a population level (see section 2.5, Chapter 2). In IID1, stool samples were collected from people who developed symptoms of diarrhoea and/or vomiting, and age and sex matched asymptomatic controls. The most suitable comparison group to contextualise my findings are the asymptomatic controls.

Table 3.9 presents the results of the DS study and the results from the IID1 study for people aged 15-74 years using conventional microbiology methods (see section 2.5 for details of the IID studies). Data from the two main components of the IID1 study are included: the community component and the GP component. Detection rates in the DS study were generally higher than those of asymptomatic controls in IID1, and for some pathogens, similar to IID1 symptomatic cases. However, IID1 was conducted over 20 years ago and the national incidence of enteric infections has changed somewhat since this time.[80] Furthermore, laboratory detection was primarily based on bacteriological culture, which is less sensitive than molecular PCR.[81] These issues should be taken into consideration when undertaking a direct comparison of the results. Re-analysis of archived stool samples from a subset of IID1 participants using PCR increased the identification of an aetiological agent from 53% to 75% in symptomatic cases and 19% to 42% in asymptomatic controls.[81] Among symptomatic cases aged 20 to 69 years, EAEC was detected in 6.7% (95% CI: 5.4% to 8.1%) of cases and 2.1% (95% CI: 1.4% to 3.1%) of asymptomatic controls using both conventional methods and PCR. Detection of *Campylobacter* spp. was 17.9% (95% CI: 15.9% to 20.0%) and 1.3% (95% CI: 0.7% to 2.2%) and detection of *Salmonella* spp. was 7.5% (95% CI: 6.2% to 9.0%) and 0.5% (95% CI: 0.2% to 1.1%) in cases and controls respectively.[81] These data suggest that detection of EAEC in the DS study was higher than asymptomatic controls in IID1 and comparable for *Campylobacter* spp. and *Salmonella* spp. Although prevalence data for adults were not available by sex, it is worth noting that rates of reported diarrhoea and/or vomiting in the community component were higher among women of reproductive age, although the confidence intervals did overlap with those of adult men.[79] In the GP component, rates of reported diarrhoea and/or vomiting in individuals aged 15 years or older were higher among women compared to men.[79]

The IID2 study was conducted during 2008-2009 but this study did not collect specimens from asymptomatic controls. Therefore, prevalence estimates can only be compared to people who developed symptoms of diarrhoea and/or vomiting (Table 3.10). The IID2 study prevalence estimates include people aged five years and over based on publicly available data. However, there was very little variation in the incidence of disease by age for people aged five years or over.[32]

In the IID2 study, *Shigella* spp. were not detected in any cases and the prevalence of EAEC was considerably lower than that in the DS study. *Campylobacter* spp. were detected more frequently in IID2 cases compared to the DS study. This is not surprising given that *Campylobacter* spp. are the most commonly isolated BEP in the UK and predominantly caused by the consumption of contaminated food. The higher detection rate in GP cases might reflect the pathogen causing more severe illness, thus influencing healthcare seeking behaviour.[80,254] It should be noted that one of the major differences in the incidence of enteric infections between IID1 and IID2 was the lower detection of BEPs in people aged five years and over, including EAEC and *Salmonella* spp., which might suggest that detection in asymptomatic controls, had they been included, would be low. The reduction in salmonellosis during this time was likely related to the UK-wide vaccination programme for chickens, which was introduced in 1998.[255]

Table 3.9: Detection of bacterial enteric pathogens in the Dean Street study and in participants aged 15 to 74 years in the IID1 study

Organism	DS study		IID1 asymptomatic controls				IID1 symptomatic cases			
			Community component		*GP component*		*Community component*		*GP component*	
	n/N	**% (95% CI)**	**n/N**	**% (95% CI)**	**n/N**	**% (95% CI)**	**n/N**	**% (95% CI)**	**n/N**	**% (95% CI)**
***Campylobacter* spp.**	35/2116	1.7 (1.2-2.3)	1/320	0.3 (0-1.7)	4/1194	0.3 (0-0.9)	15/427	3.5 (2.0-5.7)	281/1664	16.9 (15.1-18.8)
EAEC	104/2116	4.9 (4.0-5.9)	2/309	0.6 (0-0.2)	21/1185	1.8 (1.1-2.7)	10/414	2.4 (1.2-4.4)	86/1606	5.3 (4.3-6.6)
EPEC	36/2116	1.7 (1.2-2.3)	1/309	0.3 (0-0.8)	3/1185	0.2 (0-0.7)	1/414	0.2 (0-1.3)	2/1606	0.1 (0-0.4)
STEC	26/2116	1.2 (0.8-1.8)								
O157	-	-	0/320	0 (0-1.1)	0/1194	0 (0-0.3)	0/427	0 (0-0.9)	2/1664	0.1 (0-0.4)
non-O157	-	-	5/309	1.6 (0.5-3.7)	8/1185	0.7 (0.3-1.3)	0/414	0 (0-0.9)	3/1606	0.2 (0-0.5)
***Shigella* spp.**	16/2116	0.8 (0.4-1.2)	0/320	0 (0-1.1)	0/1194	0 (0-0.3)	1/427	0.2 (0-1.3)	20/1664	1.2 (0.7-1.9)
***Salmonella* spp.**	0/2116	0 (0-0.2)	1/320	0.3 (0-1.7)	5/1194	0.4 (0.1-1.0)	2/427	0.5 (0-1.7)	111/1664	6.7 (5.5-8.0)

IID1 data from Food Standards Agency (2000).[79] IID1 methods: Bacteriological culture for *Campylobacter* spp., *Salmonella* spp., *Shigella* spp. and STEC O157. DNA hybridisation techniques for EAEC, EPEC, STEC non-O157

Table 3.10: Detection of bacterial enteric pathogens in the Dean Street study and in participants aged five years and over in the IID2 study

Organism	DS Study		IID2 study – symptomatic cases			
			Community component		*GP component*	
	n/N	% (95% CI)	n/N	% (95% CI)	n/N	% (95% CI)
***Campylobacter* spp.**	35/2116	1.7 (1.2-2.3)	29/662	4.4 (3.0-6.2)	95/682	13.9 (11.4-16.8)
EAEC	104/2116	4.9 (4.0-5.9)	9/662	1.4 (0.6-2.6)	10/682	1.5 (0.7-2.7)
EPEC	36/2116	1.7 (1.2-2.3)	*Not included*		*Not included*	
STEC	26/2116	1.2 (0.8-1.8)				
O157			1/651	0.2 (0-0.9)	1/675	0.1 (0-0.8)
non-O157			6/661	0.9 (0.3-2.0)	6/681	0.9 (0.3-1.9)
***Shigella* spp.**	16/2116	0.8 (0.4-1.2)	0/651	0 (0-0.6)	0/675	0 (0-0.5)
***Salmonella* spp.**	0/2116	0 (0-0.2)	1/662	0.2 (0-0.8)	5/682	0.7 (0.2-1.7)

IID2 data from Tam *et al.* (2012).[32] IID2 results for PCR detection except STEC O157 which was detected by bacteriological culture

3.5.3.2 Comparison to studies conducted in other MSM populations

It is striking that the detection of a BEP was associated with a suite of higher-risk sexual behaviours strengthening the evidence that these pathogens are being transmitted sexually in this population of MSM.

Markers of higher-risk sexual behaviour including increasing partner number, a concurrent or previous bacterial STI diagnosis, and current HIV-PrEP use were strongly associated with BEP detection in HIV-negative/unknown status MSM. On the other hand, the association between higher-risk sexual behaviours and the detection of BEPs was less clear in the sub-group of MSM living with HIV. This could be because MSM living with diagnosed HIV were more likely to report higher-risk sexual behaviours generally compared to other MSM. Among MSM living with HIV, those who reported an 'interest in specific high-risk practices' were less likely to have a BEP detected. While this was an unexpected finding, it is important to note that this variable is used to identify men who are likely to engage in selected high-risk practices, some of which may not be directly related to BEP transmission. In addition, this

variable did not measure whether men had engaged in these activities recently or not, which could have resulted in confounding that I was unable to control for. Other factors that could have confounded this observed association are engagement in practices such as douching, or diarrhoeal illness caused by HIV medication, which could impact BEP detection. As the sample size for the subgroup of MSM living with HIV was small, it is also possible that this finding occurred by chance, particularly since the sensitivity analyses yielded conflicting findings.

My findings are comparable to a recent study carried out among 519 asymptomatic (defined as no diarrhoea in the past two weeks) MSM attending a SHC in Melbourne, Australia during November 2018 and February 2019 (Table 3.11).[256] The study found that the detection of at least one bacterial, viral or protozoan enteric pathogen in rectal swabs was independently associated with rimming in the past 12 months (aOR 3.32 [95% CI: 1.38 to 7.97]), and group sex in the past month (aOR 2.00 [95% CI: 1.11 to 3.60]).[256] Information on the number of sexual partners was not collected, which may have confounded the observed association with group sex. In contrast to my study, there was no evidence to suggest that the detection of enteric pathogens differed by HIV and PrEP status among MSM in the Australian study.

Table 3.11: Bacterial enteric pathogens detected in the Dean Street study and among MSM attending a Melbourne sexual health clinic

	DS study, n=2116	Melbourne SHC, n=519
***Campylobacter* spp.**	1.7 (1.2-2.3)	2.5 (1.5-4.3)
EAEC	4.9 (4.0-5.9)	*Not included*
EPEC	1.7 (1.2-2.3)	*Not included*
STEC	1.2 (0.8-1.8)	1.7 (0.8-3.3)
***Shigella* spp.**	0.8 (0.4-1.2)	1.0 (0.3-2.3)
***Salmonella* spp.**	0 (0-0.2)	0.4 (0-1.4)

DS study conducted from 20th December 2017 to 6th February 2018. Melbourne SHC data from Williamson *et al.* (2019) conducted between 1st November 2018 and 28th February 2019.[256]

3.5.3.3 *Sexual behaviour and STIs among MSM in the UK*

BEP detection was associated with a suite of higher-risk sexual behaviours, and a concurrent or previous bacterial STI diagnosis. In addition, higher-risk sexual behaviours were themselves highly correlated. These findings are consistent with the wider literature on sexual behaviour and STIs among MSM in the UK.[20,21,112,144,257] For instance, a large cross-sectional study of HIV-diagnosed MSM attending selected SHCs in the UK during 2011 and 2012 (The Antiretrovirals, Sexual Transmission Risk and Attitudes [ASTRA] study) found that those who reported condomless sex in the past three months were more likely to report a recent bacterial STI diagnosis, group sex, a higher number of sexual partners, and recreational and chemsex drug-use.[144] Similar findings were reported for a cross-sectional study conducted among MSM who were HIV-negative or undiagnosed and attending 20 SHCs during 2013 and 2014 (Attitudes to and Understanding of Risk of Acquisition of HIV [AURAH]); there were strong associations between measures of recreational and chemsex drug-use with a previous bacterial STI diagnosis, as well as other higher-risk behaviours such as reporting a higher number of new sexual partners, after adjustment for sociodemographic characteristics.[112]

MSM living with HIV and HIV-negative MSM attending SHCs for HIV-PrEP were more likely to have a BEP detected compared to HIV-negative MSM not taking HIV-PrEP. Evidence from community-based and online surveys suggests that MSM living with HIV generally report higher-risk sexual behaviours compared to other MSM,[20,257,258] and bacterial STI rates are also higher among MSM living with HIV compared to MSM who are HIV-negative or of unknown HIV status.[143] The association between HIV-PrEP use and BEP detection in my study is consistent with reported higher-risk sexual behaviours that facilitate STI transmission among MSM taking HIV-PrEP.[240,259]

In my study, there was no evidence of an association between BEP detection and socio-demographic factors including age, ethnic group, region of birth and IMD quintile. The existing literature on the association between socio-demographic characteristics, sexual behaviour and STIs in MSM is mixed, and varies depending on the type of sexual behaviour. Some community and SHC-based cross-sectional surveys have reported associations between measures of sexual behaviour or STIs and age. For instance, a behavioural survey of MSM

who were HIV-negative or of unknown status attending five SHCs between 2012 and 2013 found that those aged 35 to 49 years reported a higher number of sexual partners in the past three months compared to those aged 15 to 24 years, but there was no association between age group and a bacterial STI diagnosis in the same study.[260] In addition, the AURAH study reported that the prevalence of recreational and chemsex drug-use was higher among those under 45 years old,[112] and a similar finding was reported from an analysis of the two most recent London Gay Men's Sexual Health Surveys (GMSHS, community-based surveys of MSM conducted in 2013 and 2016), particularly in those aged 25 to 34 years.[257] Among HIV-diagnosed MSM in the ASTRA study, the prevalence of condomless sex and recreational drug-use was highest in those under 30 years of age and declined with increasing age group.[114,144] Consistent with the above, a 2010 London clinic-based study of HIV-diagnosed MSM also reported that men under 35 years of age were more likely to report an STI diagnosis in the past year compared to older age groups.[261] There is also some evidence that higher-risk sexual behaviour, STIs and HIV are associated with ethnicity in MSM. For instance, the prevalence of chemsex drug-use was higher among white MSM in the AURAH study, although this was of borderline significance.[112] By contrast, there was no evidence for an association between chemsex drug-use and ethnicity in an analysis of data from the two most recent GMSHS. [257] Serial cross-sectional data from the GMSHS from 2000 to 2013 found that among MSM living with HIV (diagnosed and undiagnosed), those of black ethnicity were more likely to report condomless sex and be at risk of transmitting HIV compared to those of white ethnicity.[20] In addition, analyses of STI surveillance data in England have reported that black and ethnic minority MSM are more likely to be diagnosed with a bacterial STI or HIV compared to white British MSM.[240,262] Markers of socio-economic status have also shown inconsistent associations with higher-risk sexual behaviours in MSM. In the AURAH study for example, reporting a lower level of educational attainment and financial hardship were associated with chemsex drug-use.[112] However, these factors were not associated with chemsex drug-use an analysis of data from the two most recent GMSHS.[257]

3.5.3.4 Detection of mphA

In my study, *mphA* detection indicates that this gene was present in the extracted DNA. However, it is not possible to determine whether the gene was present within a detected

enteric pathogen (those included in the study), within an undetected pathogenic organism or within other commensal gut microbes. AMR can be caused by chromosomal mutations or by the acquisition of new genes by horizontal gene transfer, that is, the movement of genetic material between organisms.[24] Horizontal gene transfer of AMR genes is often mediated by plasmids, and these can move between different species and genera of bacteria when they make contact, including bacterial cells within the host microbiota.[263,264] For example, during infection with a BEP carrying a plasmid-mediated AMR gene, the plasmid may move into a commensal bacterial cell within the gut microflora. In such a scenario, the gut microflora then acts a reservoir for AMR genes, which can then be transferred to additional pathogens if the individual acquires a subsequent infection.

There are limited data on the prevalence of *mphA* in the human gut microflora and it is expected that prevalence will vary between populations and geographical areas due to differences in antimicrobial use and infection control strategies.[265] Most studies on antimicrobial resistance have focused on detecting *mphA* in specific microorganisms isolated from the gut, usually *E. coli*.[263] In a study of 259 diarrhoeagenic and 84 commensal *E. coli* isolates from children under the age of five years in Lima, Peru, *mphA* was detected in 15.1% and 16.7%, respectively.[266] Nguyen *et al.* (2009) collected 190 *E. coli* isolates from individuals with different levels of antimicrobial exposure and a spectrum of clinical illness (healthy to hospitalised) across five countries. *MphA* was primarily detected in individuals who had received antimicrobials or who had been admitted to hospital. The detection of *mphA* in isolates from people considered to have low exposure to antimicrobials, including healthy female nurses as well as children and adults in remote settings, was low (0-4%).[238]

This study found a strong association between BEP detection and the presence of *mphA*, and *mphA* detection was associated with a bacterial STI in the past year, but only among individuals who had a BEP detected. Genomic studies on MSM-associated *Shigella* spp. have postulated that azithromycin resistance is related to off-target effects for the treatment of bacterial STIs[22] and the results from my study support this hypothesis. Given the small number of specimens where *Shigella* spp. were detected, the findings suggest that azithromycin exposure selects for resistance in a wider range of gut organisms.

3.5.4 Clinical and public health implications

My study provides evidence that the burden of BEPs among some MSM is considerable. At the outset of this study, there were only limited data on the burden of enteric pathogens in MSM, particularly for anything other than the shigellae.

The findings provide insights into the dynamics of infection transmission in MSM. When re-considering the three components of R_0 (see section 2.7.7), my data suggest that a high partner turnover in MSM might be an important driver of BEP incidence at a population level. In addition, most of the men in the study did not report any symptoms of gastroenteritis, which might suggest that asymptomatic carriage of BEPs is sustaining transmission in this population. Further understanding of the duration of asymptomatic carriage and the probability of infection transmission between an infected and a susceptible individual are needed. Furthermore, the clinical and public health implications of asymptomatic carriage are not clear. At present, infection control for sexual transmission of enteric pathogens consists of raising awareness, practicing good hygiene and abstaining from sex. If asymptomatic carriage acts a reservoir for maintaining enteric pathogen transmission in MSM, then this could represent a significant barrier to effective control, since asymptomatic screening is not recommended.

This study has suggested that there are groups of MSM who are more likely to acquire BEPs and this might allow better targeting of interventions and provide further opportunities for exploring how these infections are being transmitted, particularly among MSM living with HIV.

This study has also added to the evidence base about AMR in MSM and raises the question, 'why is the detection of *mphA* particularly high among men with a BEP?' This could reflect the fact that BEPs are themselves resistant to azithromycin, although I was unable to exclude the possibility that *mphA* was present in other organisms. The findings also provide evidence of the consequences of antimicrobial use in MSM who have behavioural risk factors for both STIs and BEPs. These MSM are exposed to high levels of antimicrobials, which may select for resistance in both pathogenic and non-pathogenic gut micro-organisms, highlighting the value of understanding behavioural risk profiles and considering off-target effects of antimicrobials in the MSM population. My findings emphasise the need to develop a holistic approach for

enteric infection prevention and management that considers the potential long-term consequences of antimicrobial treatment for STIs and other pathogens in this population.

3.5.5 Summary

This study aimed to generate estimates of prevalence for BEPs among MSM attending a SHC in central London, and to explore the socio-demographic, clinical and behavioural factors associated with prevalent infection. In summary:

- Nearly one in 10 MSM in this study had a BEP detected and most had no symptoms of gastroenteritis.
- While comparisons with existing studies are limited, the data from asymptomatic people from the IID1 study suggest that MSM may have higher levels of BEPs than might be expected in the general population.
- The detection of a BEP was associated with a suite of higher-risk sexual behaviours, strengthening the evidence that these pathogens are being transmitted sexually in this population and providing insights about the transmission dynamics.
- Compared to HIV-negative men who were not taking PrEP, those who were HIV negative and taking PrEP, and those who were living with HIV were more likely to have a BEP detected.
- Among individuals who had a BEP detected, the detection of *mphA* was associated with a previous bacterial STI, which may have implications for antimicrobial treatment guidelines.

Chapter 4: Use of molecular epidemiology to understand the distribution and genetic diversity of *Shigella flexneri* in MSM

When combined with epidemiological information on time, place and person, whole genome sequencing (WGS) data have the potential to provide novel insights into the transmission dynamics of enteric pathogens, which can be used to support national surveillance and control. In this chapter, I use WGS and epidemiological data to identify and describe the transmission of one important enteric pathogen in sexual networks of MSM, *S. flexneri*, and how this overlaps with other modes of faecal-oral transmission, with the aim of informing appropriate targeting of interventions.

4.1 Introduction

4.1.1 Public health surveillance of *Shigella* spp. in England

The Gastrointestinal Bacteria Reference Unit (GBRU) at PHE provides national microbiological reference services for a range of bacterial gastrointestinal pathogens and specialist testing for clinical, food, water and environmental samples. As well as providing advice on clinical diagnoses and management, the GBRU works at a local, national and international level to improve the detection and characterisation of pathogens and undertakes research into the genetic diversity of pathogens.[59]

Diagnostic hospital laboratories in England and Wales are requested to send pure culture of presumptive *Shigella* spp. to the GBRU for species identification and typing. Traditionally, isolates from primary culture were identified by colony appearance, using standard biochemical tests and serology performed phenotypically using agglutination tests with strain-specific antisera.[267] However, molecular typing methods now enable more precise characterisation and discrimination between strains, although with differing accuracy and discriminatory power. Molecular typing techniques for *Shigella* spp. include i) real-time polymerase chain reaction (PCR) for serotyping *S. flexneri* by targeting 10 genes specific to *S. flexneri* serotypes;[268] ii) pulsed-field gel electrophoresis (PFGE) of large genomic products generated through restriction enzyme cleavage of pathogen DNA; iii) multi-locus variable

number tandem repeat analysis (MLVA) which amplifies short sequences of repetitive DNA in the genome by PCR and separates the resultant products by gel electrophoresis; iv) multi-locus sequence typing (MLST) which uses PCR to amplify and subsequently sequence DNA from internal fragments of seven chromosomal house-keeping genes to assign a sequence type; and v) WGS to determine the entire DNA sequence of the pathogen genome.[267]

Since July 2015, WGS has been performed on all cultured isolates referred to the GBRU and is now the primary method of molecular typing.[45,269] Next generation sequencing (NGS) platforms such as Illumina (the sequencing platform used by the GBRU) have revolutionised diagnostic and public health microbiology, allowing faster sequencing at reduced costs. There are four key steps to Illumina NGS sequencing, described below and in Figure 4.1:[270]

Text removed for copyright reasons

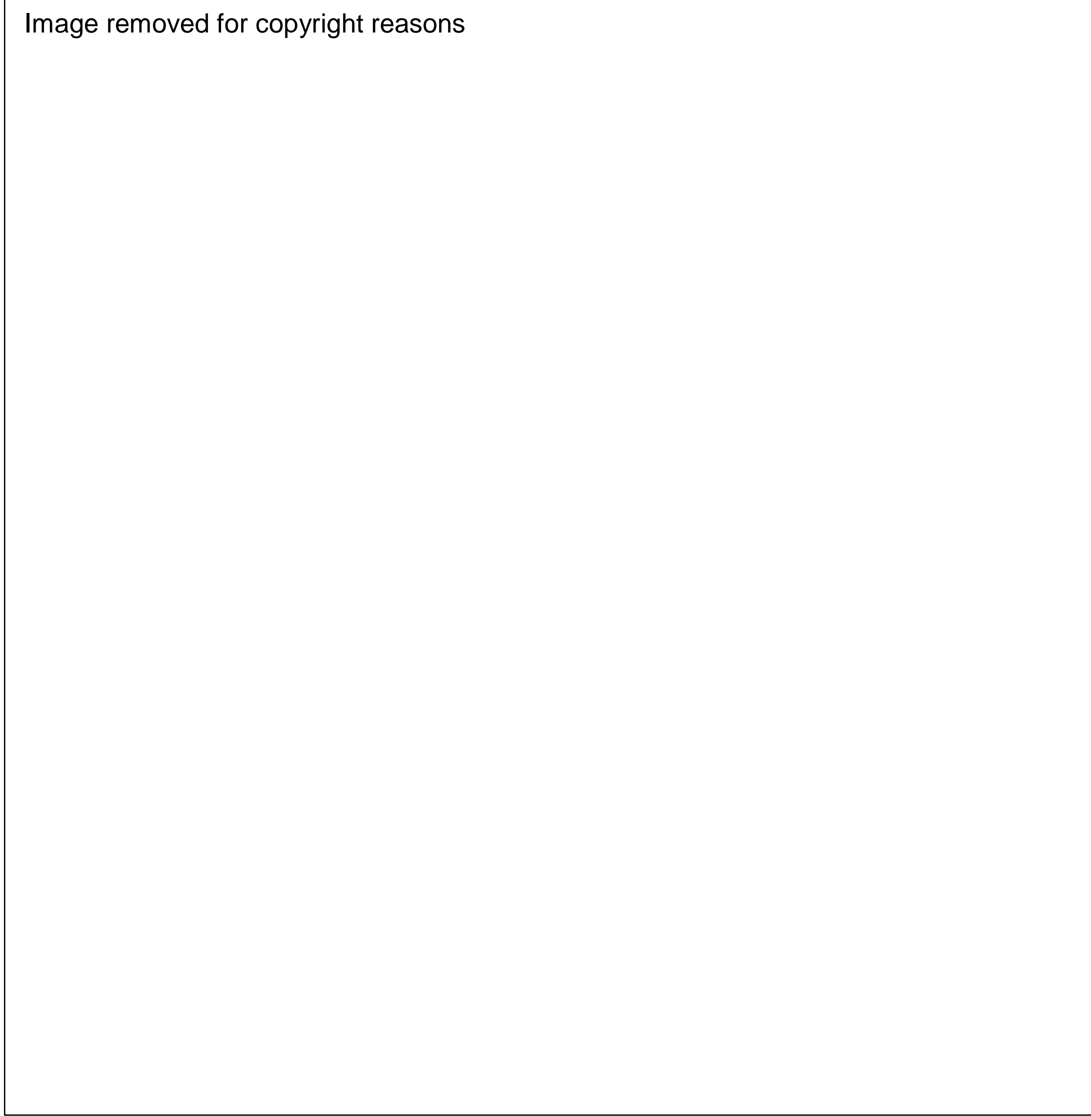


Image removed for copyright reasons

Figure 4.1: Illustration of the four steps in the Illumina whole genome sequencing process

Source: An Introduction to Next-Generation Sequencing Technology, Illumina[270]

The main advantage of WGS in comparison to other typing methods is that the entire genome of isolates can be compared in a single step.[271] This provides a higher level of discrimination that can increase the ability to distinguish between isolates. The genetic distance between isolates can be compared to infer their relatedness because isolates with similar genetic sequences are more likely to share a common ancestor, with the degree of similarity proportional to the time since divergence.[272] PHE has implemented high-throughput real-time

sequencing which draws on the power of SNP typing to support infectious disease surveillance and outbreak investigations.

4.1.2 Generation of whole genome sequencing data at PHE

4.1.2.1 DNA extraction and sequencing

The GBRU pathway from genomic DNA extraction to WGS analysis has been described in detail previously.[45,269,273] Upon arrival at the GBRU, suspected *Shigella* spp. isolates are inoculated in 1.5ml of nutrient broth and incubated overnight at 37°C. 700µl of overnight broth is added to a single well of a sterile Thermo Scientific 96 square well storage plate (deep well), followed by lysing, DNA purification and heat inactivation as described in Chapter 3, section 3.3.11.1. The plates are then transferred into the QIASymphony DNA extraction platform (Qiagen) and the genomic DNA is quantified using the Glomax (Promega).

DNA is fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation Kits (Illumina) and sequenced using the Illumina HiSeq 2500 platform.[269,274] For each sample, the output from the Illumina sequencing process are short raw DNA sequences, known as reads. These reads are assessed for quality using the Phred score, which estimates the probability of a base being incorrect. Bases with a Phred score below 30 (error probability of 1 in 1000) are removed from the trailing ends using Trimmomatic.[275]

FASTQ reads from all sequences are deposited in the PHE Pathogens Bioproject (PRJNA315192) at the National Center for Biotechnology Information (NCBI) Read Archive: https://www.ncbi.nlm.nih.gov/bioproject/315192

4.1.2.2 Sequencing analysis pipeline

The sequencing analysis pipeline is broadly divided into three steps: (i) species confirmation, (ii) sequence type and serotype identification, and (iii) SNP analysis. The first two steps aim to identify the correct pathogen-specific reference genome against which SNP typing analysis is performed.

Species confirmation is performed by comparing kmers (short strings of DNA of length *k*; in this method, *k*=18) within the reads to a set of kmers found in a set of reference genomes representing various pathogenic genera. The closest percentage match is identified and provides initial confirmation of the species.[269,274,276] The sequence type is derived by aligning the newly sequenced reads to an MLST database of reference alleles using Bowtie 2 alignment software,[277] and assigning the most likely allele at each locus to create the MLST profile (i.e. a series of seven integers that correspond to the alleles at seven different chromosomal house-keeping loci). Confidence quality metrics are reported to assess the quality of the allele assignment (i.e. the proportion of reads mapped to the reference sequence and the proportion of bases that are the same). Sequence types are subsequently grouped into clonal complexes (CC) based on their similarity to a central allelic profile. Serotyping of *Shigella* spp. is based on the structure of the O-antigen, a major component of the surface lipopolysaccharide (LPS) of Gram-negative bacteria. A customised algorithm known as 'GeneFinder'[278] utilises Bowtie 2[277] to map the newly sequenced reads to a database of sequences encoding the O-antigen synthesis and modification genes.[45,268,274] Only predictions of serotype that match to a reference gene sequence at >80% nucleotide identity over >80% of length are accepted.

For SNP typing analysis, reads are mapped to an appropriate reference strain genome using Burrows Wheeler Alignment-Maximal Exact Match (BWA-MEM).[279] The resulting Sequence Alignment Maps (SAM) are sorted and indexed to produce Binary Alignment Maps (BAM) using Samtools for faster computer processing.[280] High quality variant positions are identified using the Genome Analysis Toolkit (GATK v2.6) in unified genotyper mode[281] based on the following conditions: the specific variant is the same in over 90% of the reads, the mapping quality score is over 30 (i.e. the Phred-scored probability of the alignment being incorrect) and the sequencing depth is over 10 (i.e. the number of reads that align to a specific reference base position). These high-quality variant positions (and the ignored positions) are extracted and stored in SnapperDB, an in-house database containing a set of tools to store and analyse sequencing data from bacterial isolates.[273] As new strains are sequenced, they are added to the database and compared to the existing strains to create a distance matrix of all pairwise SNP differences.[273] The database can then be queried to output high quality variant positions

for phylogenetic analyses.[273,276] Hierarchical single linkage clustering is performed on the pairwise SNP distance matrix at various distance thresholds (250, 100, 50, 25, 10, 5 and 0 SNPs). This process results in a seven-digit 'SNP address', where each number represents clusters of isolates at each SNP distance threshold, thereby providing a nomenclature for describing the population structure based on clonal groups. Isolates are grouped into clusters if the pairwise SNP distance is within the threshold (e.g. within 10-SNPs) to at least one other isolate. Cluster detection for public health surveillance of *S. flexneri* is explored further in Chapter 5 of this thesis.

Once the sequence data have been generated, additional characterisation takes place. For example, antimicrobial resistance (AMR) determinants are detected using 'GeneFinder' (https://github.com/phe-bioinformatics/gene_finder).[278] Bowtie 2[277] is used to map newly sequenced reads to a database of reference sequences followed by SAMtools to create BAM files.[280] Genes are defined as present if they represent 100% of the reference sequence, with greater than 90% nucleotide identity.

4.1.3 Utility of WGS for infectious disease surveillance and control

Prior to the routine implementation of WGS, gastrointestinal infection outbreaks in England were investigated using both WGS and traditional typing methods for comparison purposes. These investigations showed that when combined with epidemiological data on time, place and person, SNP typing provides a high level of resolution for identifying linked cases and improving case ascertainment during outbreaks.[282-286] Two examples are provided in Box 4.1. These studies showed that the level of discrimination available through WGS enables the implementation of targeted and appropriate public health responses. WGS data have also been used to trace a multi-national outbreak of *Salmonella enterica* serovar Enteritidis [287] and to identify clusters of salmonellosis cases not detected using traditional typing methods (serotyping or phage typing).[288]

Box 4.1: Examples of the use WGS for improving case ascertainment and identifying linked cases during outbreak investigations

Example 1. An outbreak of Shiga toxin-producing *E. coli* (STEC) O157, South West England, 2014[285]

WGS of isolates from humans and cattle, combined with epidemiological investigations, identified an outbreak of STEC O157 linked to the consumption of unpasteurised cows' drinking milk produced at a dairy farm in the South West of England. In this outbreak, MLVA typing (the typing scheme in place prior to the implementation of WGS) correctly identified a link between four primary cases and one secondary case. The MLVA profile was microbiologically linked to STEC isolated from cattle on the farm. An additional nine cases were subsequently reported with the same or closely related MLVA profile. However, there was uncertainty as to whether these cases were linked to the outbreak as no epidemiological link to the consumption of unpasteurised cows' drinking milk could be found. WGS improved case ascertainment and confirmed that four of the additional nine cases were linked to the outbreak that was caused by a highly pathogenic strain of STEC. The remaining five unlinked cases were not implicated in the outbreak but were part of a wider cluster of cases that were associated with living in, or recent travel to, the South West of England.

Example 2. An outbreak of *S. sonnei* in the London Orthodox Jewish Community, 2014[284]

WGS was used to investigate an increase in cases of *S. sonnei* within the Orthodox Jewish Community (OJC) in North East London. Isolates from the outbreak were contextualised using historical isolates from previous outbreaks in this community in England, and publicly available WGS data linked to members of the OJC living elsewhere (Israel, Europe and North America). WGS revealed three concurrent regional outbreaks occurring in the OJCs across the UK, caused by multiple importations from Israel. Prior to WGS, phage typing was used to distinguish between *S. sonnei* isolates (*S. sonnei* has a single somatic (O) antigen and so cannot be serotyped) but this provided low discriminatory power; over 80% of isolates submitted to the GBRU between 2007 to 2012 belonged to one of two different phage types.[45] In this study, WGS identified clusters of closely related strains and differentiated these from the background strains circulating in the OJC population throughout the UK during the same time-period.

Source: Example 1 adapted from Butcher *et al.* (2016)[285] and example 2 adapted from Rew *et al.* (2018)[284]

4.2 Study rationale

Over the past 10 years, the epidemiology of *Shigella* spp. in England has changed from being a primarily travel-associated infection, to one where non-travel associated diagnoses in adult men account for a large and increasing proportion of all laboratory diagnoses at a national level. These data are suggestive of sexual transmission in MSM, however, the lack of information on sexual identity and behaviour hinders interpretation. In fact, there is limited direct evidence of transmission through sex between men and much of our understanding derives from semi-structured interviews (n=34) conducted during a national outbreak of non-travel related *S. flexneri* serotype 3a in 2012, which suggested that specific sexual activities and drug-use behaviours predominantly among MSM living with HIV were facilitating sexual transmission.[4]

Genomic data from nationally representative sub-sets of isolates in the UK (2004 to 2014) have been used to describe both the global and regional spread of *Shigella* spp. among large networks of MSM.[22,168] However, these studies lacked information on sexual identity and behaviour, and used circumstantial routine demographic data from laboratory report forms (age, gender, and recent foreign travel history), supplemented with enhanced behavioural data for a small sub-set of isolates (n=54 out of 697 isolates across two studies), to infer that these sub-lineages were sexually transmitted within networks of MSM.[22,168]

In 2015, to address a lack of direct evidence of sexual transmission at a national level and to inform infection control measures at an individual, local and national level, PHE piloted a new questionnaire to standardise and expand the collection of exposure information on suspected cases of *S. flexneri*, *S. dysenteriae* and *S. boydii*; this information is not routinely collected for *S. sonnei* cases because the public health management of cases and contacts is different to non-*sonnei Shigella* spp.*,* which usually cause more severe illness. For the first time, the questionnaire included questions about the case's sexual identity and recent sexual behaviour. As well as ensuring that men receive appropriate advice (for example, about the importance of being tested for STIs, HIV and blood-borne viruses), these questions were added to allow PHE to better monitor sexual transmission as a risk factor for shigellosis and to inform targeted infection control efforts that seek to prevent onward transmission, either

through faecal-oral transmission in the community or sexual transmission. Concurrently, WGS was introduced as a routine procedure for all *Shigella* spp. isolates referred to the reference laboratory at PHE.[45,269] Combined, WGS and sexual behaviour data offer the opportunity to improve our understanding of *Shigella* spp. transmission within sexual networks of MSM to better inform the public health response.

In this chapter, I have used rich epidemiological data combined with WGS data of *S. flexneri* to address the following specific research questions:

1. What molecular, clinical and epidemiological characteristics are associated with the sexual transmission of *S. flexneri* in MSM and could inform targeted prevention activities?
2. What is the relationship between genotypic markers of azithromycin resistance and sexual transmission of *S. flexneri* which could inform guidelines on the use of antimicrobials?

I chose to focus the analyses on *S. flexneri*, because over 95% of case questionnaires were from people diagnosed with this species of *Shigella*. In addition, all MSM with questionnaire data were diagnosed with *S. flexneri*.

4.3 Aim and Objectives

In this study, I aimed to combine socio-demographic, sexual behaviour and clinical data from case questionnaires with routine WGS data to identify and characterise *S. flexneri* transmission through sex between men and to understand how this overlaps with non-sexual transmission within the community. The objectives were to:

1. Describe the genetic diversity of *S. flexneri* isolates submitted to the national reference laboratory and how this varies for MSM and non-MSM
2. Describe the distribution of genotypic markers of azithromycin resistance
3. Explore the epidemiological, clinical and molecular characteristics associated with *S. flexneri* in MSM and non-MSM

4. Explore the epidemiological, clinical and molecular characteristics associated with clinical severity of *S. flexneri* infections

4.4 Methods

4.4.1 *S. flexneri* WGS data

I included all *S. flexneri* isolates referred to the national reference laboratory in England between August 2015 and July 2017 in my study. I used the Gastro Data Warehouse (GDW), an isolate-level database that stores GBRU laboratory results for gastrointestinal bacteria, to extract sequencing results (isolate identifier, date of specimen, species, sequence type, serotype, AMR genes and SNP typing results) and demographic data (name, date of birth, sex, postcode of residence and foreign travel history). Duplicate isolates belonging to the same individual within a two-week period were excluded, according to standard PHE protocols.

4.4.2 Standardised shigellosis exposure questionnaire

I used data collected using a PHE standardised questionnaire for following-up cases of shigellosis that was piloted from August 2015 to March 2017 by seven Health Protection Teams (HPTs) in England (all three in London, four outside London); there are 21 HPTs in England altogether (three in London, 18 outside London). The questionnaire collected self-reported information including demographics, sexual identity for cases aged 18 years or older (heterosexual/straight, gay/lesbian, bisexual, other, don't know/refuse to answer, based on the Office for National Statistics [ONS] question on sexual identity[289]), sexual contact for adult men aged 18 years or older in the past four days (recent sexual contact [Yes/No] and if yes, was this with a man and/or a woman, or prefer not to answer), recent foreign travel (past four days), food and water consumption (past four days), clinical condition (date of onset, symptoms and hospitalisation), and risk group status (i.e. at increased risk of spreading the infection to others such as a food handler, healthcare worker or those in contact with children aged five years or under (see section 2.6, Table 2.4).[46] A copy of the questionnaire can be found in Appendix 4.1.

Paper-based questionnaires were completed by HPTs or Environmental Health Officers (EHOs) and sent to a dedicated email address at PHE. The national STI surveillance team at PHE collated questionnaire responses and entered them into an in-house database. I extracted and cleaned the questionnaire data using Stata v15.1. First, I checked for duplicate questionnaires from the same person. Second, I checked missing responses using the original paper-based questionnaires. Third, I performed consistency checks by cross-tabulating variables, for instance I checked that sexual identity was reported for adults and that recent sexual behaviour was reported for adult men. Fourth, I re-coded and grouped the data to create meaningful variables for the analyses, for instance date of birth was used to calculate age and I created a new 'age group' variable. Where possible, I verified invalid responses using the original questionnaires or by contacting the HPT staff involved in data collection, for instance, where postcode data did not match to any known UK postcodes.

4.4.3 Linkage of WGS and questionnaire data

I linked questionnaire data to WGS data extracted from GDW using a combination of first name, surname, date of birth, sex, and full postcode of residence. For matched isolates, I cross-checked the sample date recorded on the questionnaire with that recorded in GDW. This was important because some people had multiple isolates for different *Shigella* spp. episodes recorded in GDW. For isolates that did not match, I first checked whether there were any obvious typos that could have resulted in a miss-match, for instance, if patient name was spelled incorrectly in one dataset but all other identifiers matched. These miss-matches were updated as appropriate using the questionnaire data as the gold standard. Once all matches were complete, I cross-checked foreign travel data between the two datasets and updated this using the questionnaire data. All data cleaning, management and linkage were performed using Stata v15.1.

4.4.4 Linkage to national HIV surveillance data

Previous studies have suggested that HIV is a risk factor for shigellosis,[121,122] and information on HIV status could improve our understanding of the relationship between these infections in sexual networks of MSM.

PHE collects data for all people diagnosed with HIV in the UK as part of the national HIV surveillance programme.[290] The data consist of four linked sources; new HIV and AIDS diagnoses and deaths, laboratory reports of CD4 cell counts, follow-up information on clinical outcomes, treatment prescribing and co-morbidities collected annually from all NHS HIV outpatient clinics, and reports of HIV-related death from the ONS. The HIV and AIDS Reporting System (HARS) is a consultation-based disaggregate dataset, which is part pseudo-anonymised i.e. the data contain date of birth and Soundex code, but not patient name. Soundex is a coding system for names based on phonetic spelling that generates an anonymous identifier. The code consists of the first letter of the surname and three digits that represent the first three phonetic sounds in the name.[291]

HIV surveillance data are managed by the national HIV surveillance team at PHE using standardised protocols. To comply with strict information governance protocols for handling data on HIV cases, linkage to the HIV database was performed by a member of the HIV surveillance team. First, the HIV surveillance scientist modified the linked *S. flexneri* WGS and questionnaire dataset (see section 4.4.3) to minimise the risk of disclosure for people living with diagnosed HIV and to create the variables required to facilitate linkage (i.e. the case's surname was replaced with a Soundex code and first name was replaced with a first initial). Linkage to the HIV database was performed using a hierarchical matching algorithm which prioritised higher confidence matches, and was based on an algorithm that had been developed for a previous study.[121] After matching, all personal identifiers (date of birth, postcode) and pseudo-anonymised identifiers (Soundex code, first initial) were removed and the data were irreversibly anonymised with a new unique identifier prior to returning to me for further analysis. For matched cases, I included information on HIV diagnosis date, probable route of exposure to HIV, and most recent CD4 and viral load count (within three months of the *S. flexneri* episode) in my analyses.

4.4.5 Data analysis

4.4.5.1 Definitions

Table 4.1 provides a summary of the main terms used to describe the different *S. flexneri* cases analysed in this chapter.

Table 4.1: Summary of main terms and their definitions

Term	Definition
Confirmed MSM	Men who self-identified as gay or bisexual or who reported recent same-sex sexual contact on the questionnaire
Confirmed non-MSM	Adult men who self-identified as heterosexual, adult women and children under the age of 18 years as reported on the questionnaire
Confirmed other adults	Adult men who self-identified as heterosexual and adult women as reported on the questionnaire. This group is a sub-set of confirmed non-MSM.
Presumed MSM	Adult men who did not have questionnaire data but were presumed to be MSM based on phylogenetic analyses of their *S. flexneri* isolates
MSM clade	Phylogenetic clade (serotype 2a/3a) where a large proportion of *S. flexneri* isolates were confirmed to be MSM based on the questionnaire data
Travel-associated lineage	Phylogenetic lineages where a large proportion of *S. flexneri* isolates were associated with recent foreign travel

4.4.5.2 Descriptive and statistical analyses

I performed descriptive analyses of the epidemiological, molecular (including phylogenetic inferences and genotypic markers of azithromycin resistance) and clinical characteristics of confirmed MSM and confirmed non-MSM. Differences between the two groups were assessed using the Chi-squared test for comparing two proportions (for categorical variables) or Wilcoxon rank sum tests (for continuous variables). Adult men who did not provide information on sexual identity or recent sexual behaviour were not included in these statistical comparisons.

I explored epidemiological, molecular and clinical characteristics associated with severe clinical symptoms and outcomes of *S. flexneri* infection in adults using univariable and age-adjusted logistic regression. To explore these associations further, I also created two

composite outcomes of severity: (i) severe clinical symptoms defined as the presence of blood and/or mucus in stools compared to the absence of these symptoms, and (ii) clinical outcomes defined as hospital admission and/or antimicrobial use compared to the absence of these outcomes. Logistic regression is the most common statistical model used to analyse binary outcome (or dependent) variables. This model measures the association between a binary outcome variable and one or more independent variables using the odds ratio (OR).[252] I chose this modelling approach for the analysis because I was exploring the characteristics associated with several different outcome variables. With logistic regression, the magnitude of association can be directly compared across various binary outcome variables of different prevalence. Such direct comparison is not possible using Poisson regression with robust error variances, the modelling approach used in Chapter 3. I did not conduct multivariable analyses adjusting for multiple independent variables because of the smaller sample size and because many of the independent variables were highly correlated. The interpretation of regression analyses is particularly challenging when there are high levels of correlation between independent variables (see section 4.6.3.2 below).

For all clinical symptoms and outcomes, missing data were taken to indicate the absence of the specific symptom/outcome. To determine whether this was a valid assumption, I repeated the analyses excluding people with missing data on each specific symptom/outcome. All analyses were carried out using Stata v15.1.

4.4.5.3 Phylogenetic analyses

I generated phylogenetic trees using the functions provided in SnapperDB.[273] First, I generated whole genome alignments in FASTA format for a specified set of isolates. The alignments were created using the mapping tool BWA-MEM and high-quality SNPs were identified and extracted using GATK v2.6 as described in section 4.1.2.2. The whole genome alignment includes conserved and variable sites with respect to the reference genome. The FASTA file was input into Gubbins v2.0 to detect possible recombination events within the genome by assessing the SNP density at each site.[292] I used the GFF output file of recombination predictions to create pseudo-alignments of polymorphic positions in FASTA format, with recombinant regions of the genome removed. The pseudo sequences of polymorphic positions

were used to create maximum likelihood trees using RAxML v8.2.8 under the General Time Reversible model using up to 1000 bootstrap replicates.[293] The analyses were performed in two steps. First, a single representative isolate from each 10-SNP single linkage cluster was used to understand the phylogenetic context and describe the population structure and second, single linkage clusters were purposively sampled and the phylogeny reconstructed using all isolates within that cluster. The output phylogenetic tree was midpoint rooted (meaning the root of the tree was placed half-way between the two most distant specimens) and annotated with epidemiological data using Interactive tree of life (iTOL) v4.3.[294,295] iTOL enabled interactive visualisation of the trees with the epidemiological data, which are presented using different colours and symbols for the variable values.

Timed phylogenies were reconstructed using BEAST2[296] with a strict molecular clock and both constant and exponential population growth models (results from both models were compatible). Both models were run with a chain length of 10 million. A maximum clade credibility tree was reconstructed using TreeAnnotator v1.75 and annotation was performed using FigTree v1.4.3.

4.4.6 Information governance and ethics

No specific consent was required from the patients whose data were used in these analyses. PHE has authority to handle patient data for public health monitoring and infection control under section 251 of the UK National Health Service Act of 2006 (previously section 60 of the Health and Social Care Act of 2001), which was reviewed annually by the ethics and confidentiality committee of the National Information Governance Board until 2013. Since then the power of approval of public health surveillance activity has been granted directly to PHE, and is operated through the PHE Caldicott Panel.

My PHE honorary contract was finalised in May 2017 and covered all the relevant data confidentiality arrangements regarding access to national surveillance data held by PHE. Information governance advice and ethical approval for the analyses were sought from the PHE Research Support and Governance Office (RSGO). The RSGO approved the analyses as falling within public health surveillance and as such, no ethical approval was required. Thus,

in accordance with standard PHE procedure for public health surveillance, the analyses were reviewed and approved by the PHE Caldicott Panel in June 2017 (Appendix 4.2).

To ensure anonymity of people living with HIV, *S. flexneri* data linked to the HIV database were irreversibly anonymised prior to my analysis. I did not have access to the original HIV data and all matching was performed by a member of the national HIV surveillance team at PHE.

4.5 Results

4.5.1 Description of study isolates and questionnaire data

Figure 4.2 describes the number of *S. flexneri* isolates submitted to the national reference laboratory during the study period (August 2015 to July 2017) including showing the number of isolates with linked questionnaire data. In total, 1,006 *S. flexneri* isolates were reported, of which 92.0% (926) belonged to clonal complex (CC) 245 and 8% (80) belonged to CC145. Of CC245 isolates and where demographic data were reported, the majority were from males (72.2% [659/913]) and from adults aged 18 years or older (84.2% [775/920]), and less than one third reported recent foreign travel (29.4% [272/926]). Of CC145 isolates, the majority were from females (63.8% [51/80]) and from adults aged 18 years or older (68.8% [55/80]), and about half (52.5% [42/80]) were from people who reported recent foreign travel.

Linked questionnaire data collected during the pilot period (August 2015 to March 2017) were available for 190 *S. flexneri* isolates, of which 95.8% (n=182) belonged to CC245 and 4.2% (n=8) belonged to CC145 (Figure 4.2). Where questionnaire data were available, half of cases (50.0%; 95/190) represented self-confirmed MSM (see definitions section 4.4.5.1), 35.3% (n=67) were other adults (men who identified as heterosexual or adult women), 10% (n=19) were children under the age of 18 years and 4.7% (n=9) were adult men who did not provide information on sexual identity or recent behaviour. The isolates with linked questionnaire data represented 37.8% (190/503) of all isolates referred to the national reference laboratory from the HPTs that participated in the pilot of the national questionnaire (42.4% [143/337] in

London, 28.3% [47/166] outside London), and 21.9% (190/868) of all isolates nationally during the pilot period (August 2015 to March 2017).

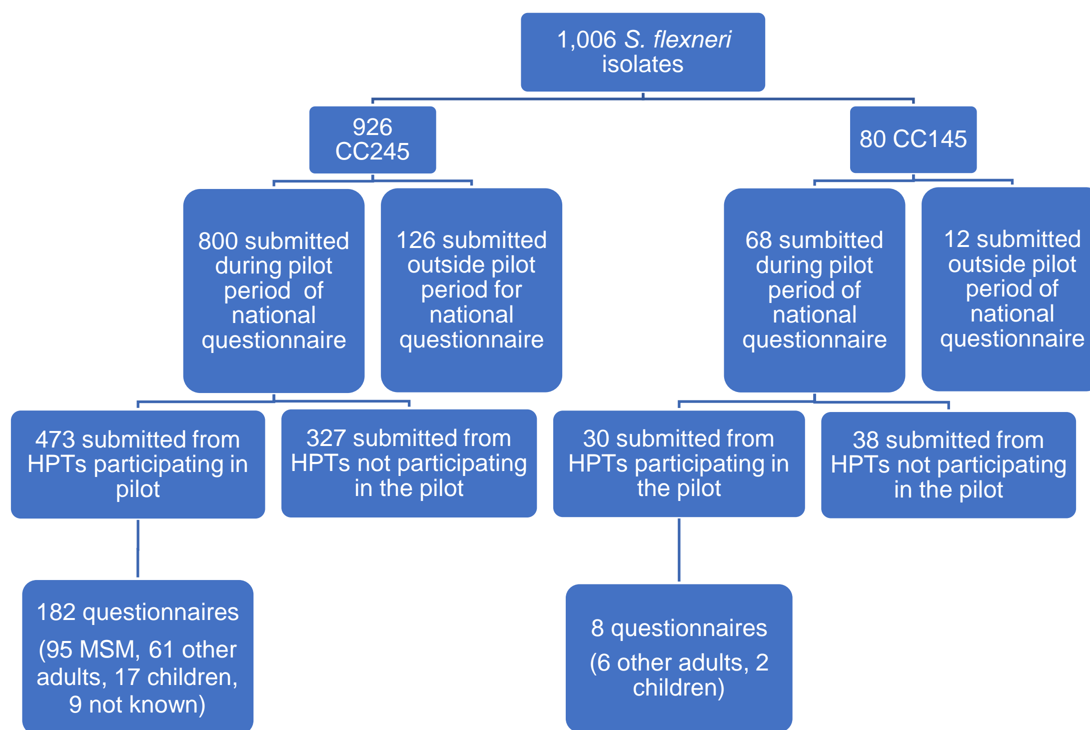


Figure 4.2: Total number of *S. flexneri* isolates and the number with linked questionnaire data

All subsequent analyses presented in this chapter include isolates and linked metadata belonging to CC245 only. This includes 926 isolates, of which 182 (19.7%) had linked questionnaire data. I chose to do this because i) all isolates from MSM belonged to CC245, ii) I only had limited questionnaire data for isolates within CC145 (4.2%; 8/190), and iii) SnapperDB, the database application used to analyse bacterial WGS data at PHE, is intended for use at a clonal complex level.

Among isolates submitted from HPT regions during the pilot period (August 2015 to March 2017), isolates with questionnaire data were more likely to be submitted from London compared to isolates without questionnaire data (75.3% vs 63.9%, p=0.010) (Table 4.2). There was no difference between isolates with and without questionnaire data from HPT regions according to age group, gender or recent foreign travel. Among all isolates included in the

study (August 2015 to July 2017), isolates with linked questionnaire data were more likely to be submitted from London compared to isolates without questionnaire data included in the study (75.3% vs 32.9%, p<0.001) (Table 4.2). There was also a higher proportion of isolates with linked questionnaire data from adults aged 25-34 years compared to those with no questionnaire data (31.3% vs 21.5%).

Table 4.2: Epidemiological characteristics of *S. flexneri* CC245 isolates with and without linked questionnaire data in England, August 2015 to July 2017

Variable	Questionnaire (n=182)		No questionnaire: pilot HPTs (N=291)		p-value (questionnaire vs no questionnaire pilot HPTs)	No questionnaire: all isolates (n=744)		p-value (questionnaire vs no questionnaire all isolates)
	N	%	N	%		N	%	
HPT								
London	137	75.3	186	63.9	0.010	245	32.9	<0.001
Non-London	45	24.7	105	36.1		499	67.1	
Gender								
Male	140	76.9	217	75.9	0.795	519	71.0	0.111
Female	42	23.1	69	24.1		212	29.0	
Not known	0	0	5			13		
Age group								
<18	17	9.3	39	13.5	0.492	128	17.3	0.016
18-24	15	8.2	25	8.6		67	9.1	
25-34	57	31.3	75	25.9		159	21.5	
35-44	36	19.8	66	22.8		141	19.1	
45+	57	31.3	85	29.3		243	32.9	
Not known	0	0	1			6		
Recent foreign travel*								
No/unknown	146	80.2	226	77.7	0.509	547	73.5	0.062
Yes	36	19.8	65	22.3		197	26.5	

N=926 isolates overall. 473 isolates submitted by HPT pilot regions (August 2015 to March 2017). *Recent foreign travel based on data reported on GDW here to enable comparison between two groups. Foreign travel data in GDW are underreported. P values calculated using Chi-squared test.

4.5.2 Describing the phylogeny

4.5.2.1 Genetic diversity of cases

The phylogenetic analysis of 926 CC245 isolates revealed two domestically circulating clades first described by Baker *et al.* (2015, 2018)[22,168] that were considered to be associated with transmission in MSM (herein referred to as 'MSM clades') (Figure 4.3); one phylogenetic clade within *S. flexneri* phylogenetic group (PG) 3 (median pairwise SNP distance 21; minimum 0, maximum 47) serotype 2a, and a second phylogenetic clade within PG1 (median pairwise

SNP distance 37, minimum 0, maximum 165) serotype 3a. These two clades accounted for 43.0% of all isolates included in my study (33.8% [313/926] to PG3 and 9.2% [85/926] to PG1). 97.1% (300/309) of isolates in the PG3 MSM clade were from adult men (aged 18 years or older), 1.9% (6/309) were from adult women and 1.0% (3/309) were from children. Gender was not recorded for four isolates belonging to adults. Among isolates in the PG1 MSM clade, 98.8% (84/85) of isolates were from adult men and one isolate was from a child.

Overlaying the phylogenetic tree with questionnaire data revealed that 74.7% (71/95) of all isolates from MSM belonged to the domestically circulating MSM clade within PG3 and 13.7% (13/95) to the MSM clade within PG1 (Figure 4.3, Table 4.3). The remaining 11.6% (11/95) of isolates from MSM were dispersed throughout the phylogeny within travel-associated lineages. A small number of men self-identifying as heterosexual (6/27) or not providing sexual identity or behaviour data (7/9) had isolates that were phylogenetically located within the two dominant domestically circulating MSM clades. Among men reporting heterosexual identity, three reported recent sexual contact with a woman, one reported no recent sexual contact, and two did not provide this information.

4.5.2.2 Distribution of genotypic markers of azithromycin resistance

Overall, 40.0% (370/926) of all *S. flexneri* isolates included in this study harboured genotypic markers of azithromycin resistance (*mphA* and/or *ermB*), of which 89.2% (330/370) were phylogenetically located within the domestically circulating MSM clades described above and 10.8% (40/370) were dispersed on discrete branches throughout the phylogenetic tree ($p<0.001$) (Figure 4.3). Of the 40 isolates with genotypic markers of azithromycin resistance located within travel-associated lineages, 31 were from adult men, four were from adult women and five were from children under the age of 18 years. Six of these people had questionnaire data; five were MSM and one was an adult man who did not disclose their sexual identity nor the gender of their recent sex partner.

Among isolates with linked questionnaire data, those from confirmed MSM were more likely to harbour *mphA* and/or *ermB* compared to isolates from confirmed non-MSM (Figure 4.3, Table 4.3): 83.2% (79/95) of isolates from confirmed MSM harboured genotypic markers of resistance compared to only 7.7% (6/78) of isolates from confirmed non-MSM cases

(p<0.001), and all of the latter were from heterosexual-identifying men whose isolates were phylogenetically located within the domestically circulating MSM clades.

4.5.3 Overlap between *S. flexneri* and HIV

I used data linkage to ascertain the HIV status of cases infected with *S. flexneri* and found that 18.7% (173/926) of CC245 isolates included in this study were from people living with HIV; 170 were from adult men and three were from adult women (Table 4.4). The results from the hierarchical data linkage procedure can be found in Appendix 4.3. Among those living with HIV, 86.1% (149/173) of *S. flexneri* isolates belonged to the two domestically circulating MSM clades. The probable route of HIV exposure for *S. flexneri* cases living with HIV and whose isolates belonged to the MSM clades was sex between men for 90.6% (135/149), heterosexual contact for 4.7% (7/149), injecting drug use for 1.3% (2/149) and unknown for 3.4% (5/149). Among the MSM clades, 38.7% (121/313) of *S. flexneri* isolates from PG3 and 32.9% (28/85) of isolates from PG1 were from people living with HIV, of whom 88.5% (131/148) were diagnosed with HIV more than six weeks before their *S. flexneri* diagnosis.

4.5.4 Epidemiological and clinical characteristics of *S. flexneri* cases

Among people with questionnaire data, confirmed MSM were less likely to report recent foreign travel (past 4 days) compared to confirmed non-MSM (12.6% [12/95] vs 78.2% [61/78], p<0.001) (Figure 4.3; Table 4.3). Most confirmed MSM who reported recent foreign travel had travelled to Europe (83.3%; 10/12). In contrast, most confirmed non-MSM who reported recent foreign travel had visited regions outside of Europe (98.4%; 60/61), predominantly South Asia (44.3%; 27/61) or sub-Saharan Africa (31.1%; 19/61). Most confirmed MSM reported recent sex with a same-sex partner in the four days prior to symptom onset (72.3%; 68/94), of whom 13.2% (9/68) had also travelled, mainly to Europe (n=7). Among confirmed MSM without recent sexual contact (27.7%; 26/94), only two had recently travelled and this was to Europe. Confirmed MSM were also more likely to be of white ethnicity compared to confirmed non-MSM (79.6% [70/94] vs 34.3% [24/70], p<0.001). One in five confirmed MSM (19/90; 21.1%) and confirmed other adults (11/55; 20.0%) were in a group considered to be at increased risk of spreading infection to others i.e. their occupation was categorised as clinical, social care,

nursery worker or food handler. Among confirmed MSM who were living with HIV, nearly all (95.9%; 47/49) were diagnosed prior to their *S. flexneri* infection; where data were available, 80.8% (21/26) had an undetectable viral load (≤50 c/ml) and 81.2% (18/21) had a CD4 count greater than 350 cells/mm^3.

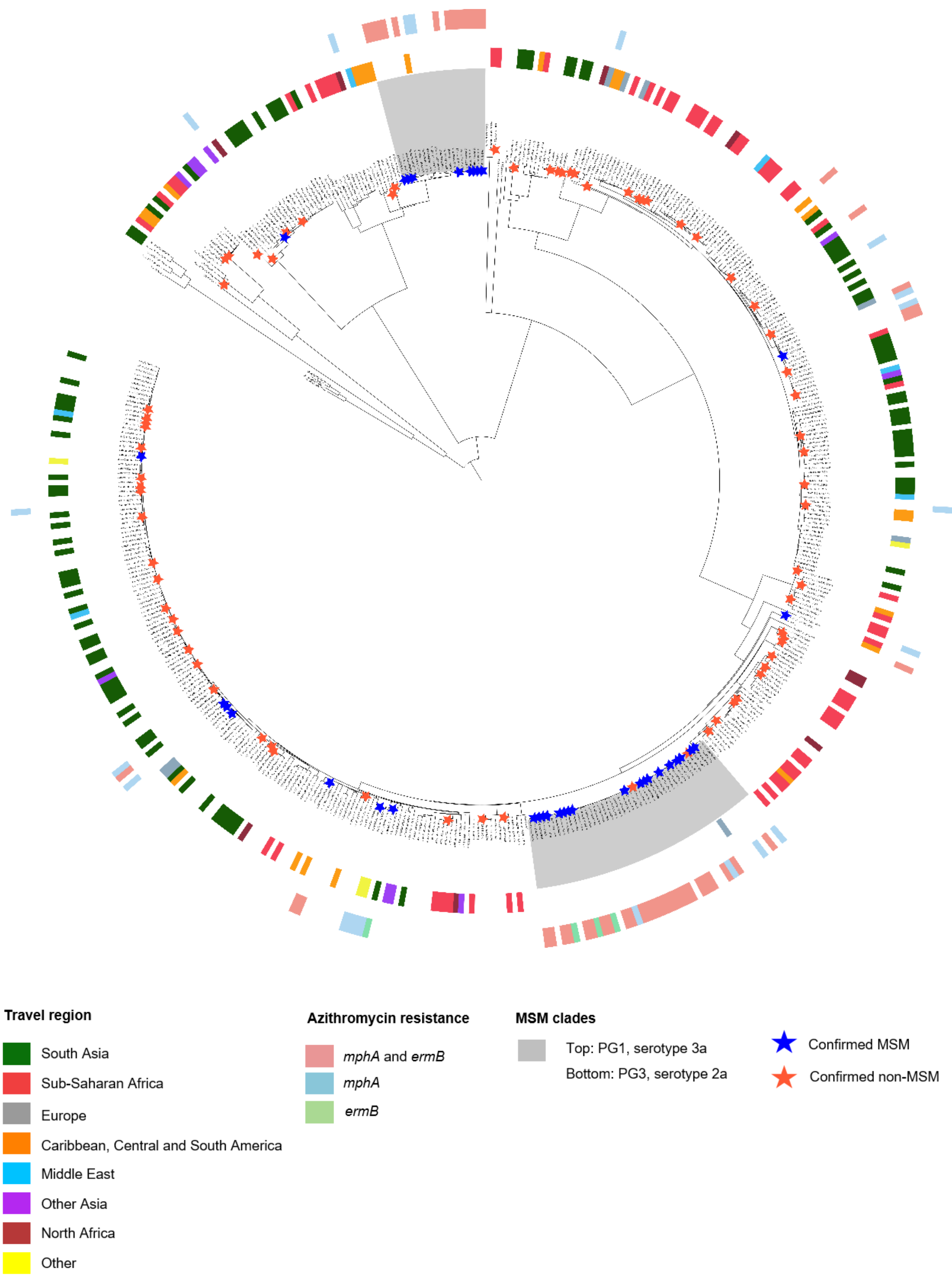


Figure 4.3: Phylogeny of *S. flexneri* CC245 isolates, August 2015 to July 2017

Mid-point rooted maximum likelihood phylogenetic tree showing a single representative isolate from each 10-SNP single linkage cluster (N=474) for CC245 during the study period (N=926 isolates) and seven reference strains for each phylogenetic group.[297] The number of isolates represented by each tip ranges from 1 to 240. Region of travel (inner track) and genotypic markers of azithromycin resistance (outer track) are shown as coloured tracks on the outside of the tree. Isolates belonging to confirmed MSM and confirmed non-MSM (identified through questionnaire data) are shown as stars on the branches, and MSM-associated clades are highlighted in grey.

Table 4.3: Selected epidemiological, clinical and molecular characteristics for cases with questionnaire data

Characteristic	MSM N=95	Other adults N=61	Children N=17	Not known N=9
Gender				
Male	95 (100)	27 (44.3)	9 (52.9)	9 (100)
Female	0	34 (55.7)	8 (47.1)	0
Age group				
<18	0	0	17 (100)	0
18-24	8 (8.4)	6 (9.8)	0	1 (11.1)
25-34	33 (34.7)	24 (39.3)	0	0
35-44	25 (26.3)	8 (13.1)	0	3 (33.3)
45-64	28 (29.5)	17 (27.9)	0	5 (55.6)
65+	1 (1.1)	6 (9.8)	0	0
Ethnic group				
White	70 (79.6)	22 (40.7)	2 (12.5)	5 (100)
Asian or Asian British	3 (3.4)	18 (33.3)	11 (68.8)	0
Other	15 (17.1)	14 (25.9)	3 (18.8)	0
Not specified	7	7	1	4
Recent foreign travel history (past 4 days)				
South Asia	0	18 (29.5)	9 (52.9)	0
Sub-Saharan Africa	0	17 (27.9)	2 (11.8)	0
Europe	10 (10.5)	1 (1.6)	1 (5.9)	1 (11.1)
Caribbean, Central and South America	1 (1.1)	5 (8.2)	0	0
Middle East	1 (1.1)	2 (3.3)	1 (5.9)	1 (11.1)
North Africa	0	3 (4.9)	1 (5.9)	0
Other Asia	0	1 (1.6)	0	0
No/Not specified±	83 (87.4)	14 (23.0)	3 (17.7)	7 (77.8)
Sexual identity (n=165)				
Gay man	92 (98.9)	0	-	0
Bisexual man	1 (1.1)	0	-	0
Heterosexual man	0	27 (50.0)	-	0
Heterosexual woman	0	27 (50.0)	-	0
Not specified	2*	7**	-	9
Recent sexual contact (past 4 days) (n=131)				
Yes – with man	68 (72.3)	0	-	0
Yes – with woman	0	12 (50.0)	-	0
Yes – gender of partner not disclosed	0	0	-	2 (33.3)
No	26 (27.7)	12 (50.0)	-	4 (66.7)
Not specified	1	3	-	3
IMD quintile				
1 (Most deprived)	40 (42.1)	20 (32.8)	5 (29.4)	1 (12.5)
2	38 (40.0)	20 (32.8)	8 (47.1)	6 (75.0)
3	9 (9.5)	14 (23.0)	2 (11.8)	0
4	3 (3.2)	4 (6.6)	2 (11.8)	1 (12.5)
5 (Least deprived)	5 (5.3)	3 (4.9)	0	0
Not specified	0	0	0	1

Characteristic	MSM N=95	Other adults N=61	Children N=17	Not known N=9
Occupation				
School/nursery child	0	0	17 (100)	0
Health care¥	8 (8.9)	1 (1.8)	0	1 (12.5)
Social care/nursery worker¥	3 (3.4)	4 (7.3)	0	0
Food handler/catering ¥	8 (8.9)	6 (10.9)	0	0
Fitness/gym worker	0	2 (3.6)	0	0
Travel industry	2 (2.2)	0	0	0
Other	55 (61.1)	34 (61.8)	0	6 (75.0)
Not working/retired	14 (15.6)	8 (13.1)	0	1 (12.5)
Not specified	5	6	0	1
Serotype				
2a	73 (86.9)	27 (45.8)	10 (62.5)	5 (55.6)
Other	11 (13.1)	32 (54.2)	6 (37.5)	4 (44.4)
Not specified	11	2	1	0
Phylogenetic lineage				
MSM clade (PG3, serotype 2a)	71 (74.7)	5 (8.2)	0	5 (55.6)
MSM clade (PG1, serotype 3a)	13 (13.7)	1 (1.6)	0	2 (22.2)
Travel-associated lineage	11 (11.6)	55 (90.2)	17 (100)	2 (22.2)
Genotypic markers of azithromycin resistance				
mphA and *ermB*	71 (74.7)	5 (8.2)	0	5 (55.6)
mphA only	4 (4.2)	1 (1.6)	0	1 (11.1)
ermB only	4 (4.2)	0	0	0
None	16 (16.8)	55 (90.2)	17 (100)	3 (33.3)
HIV status at *S. flexneri* diagnosis				
HIV-negative/unknown status	45 (47.9)	58 (95.1)	17 (100)	7 (77.8)
HIV diagnosed more than 6 weeks previously	45 (47.9)	3 (4.9)	0	2 (22.2)
HIV diagnosed within previous 6 weeks	2 (3.2)	0	0	0
HIV diagnosed within 6 weeks after	1 (1.1)	0	0	0
HIV diagnosed more than 6 weeks after	1	0	0	0
HIV diagnosis date not known¥	1	0	0	0

N=182 unless otherwise specified; denominator for sexual identity includes adults aged 18 years or older and the denominator for recent sexual contact (past four days prior to symptoms) includes adult men aged 18 years or older. Missing data excluded from percentage calculations except for recent foreign travel history. ±Recent foreign travel (past four days prior to symptoms) as recorded on questionnaire; data missing for two cases. *Sexual identity not specified for two men, but recent same-sex sexual contact reported. **Sexual identity not specified for seven adult women. ¥Occupation indicates the patient belongs to a recognised risk group and poses an increased risk of spreading their infection to others. ¥Living with HIV but HIV diagnosis date not known. IMD: Index of Multiple Deprivation.

Table 4.4: Epidemiological and molecular characteristics of people living with HIV

Characteristic	MSM clade (N=149)	Travel-associated lineage (N=24)
Gender		
Male	149 (100)	21 (87.5)
Female	0	3 (12.5)
Age group		
18-24	8 (5.4)	0
25-34	40 (26.9)	3 (12.5)
35-44	57 (38.3)	5 (20.8)
45-64	44 (29.5)	16 (66.7)
Recent foreign travel		
Yes	7 (4.7)	7 (29.2)
No/not specified	142 (95.3)	17 (70.8)
Sexual identity (n=55)		
Gay man	44 (93.6)	5 (83.3)
Bisexual man	1 (2.1)	0
Heterosexual man	2 (4.3)	1 (16.7)
Not specified	2	0
Probable route of exposure to HIV		
Sex between men	135 (93.8)	16 (76.2)
Injecting drug use	2 (1.4)	0
Heterosexual contact – man	7 (4.9)	3 (14.3)
Heterosexual contact – woman	0	2 (9.5)
Not known	5	3
HIV status at time of *S. flexneri* diagnosis		
HIV diagnosed more than 6 weeks previously	131 (88.5)	21 (87.5)
HIV diagnosed within previous 6 weeks	4 (2.7)	1 (4.2)
HIV diagnosed within 6 weeks after	3 (2.0)	2 (8.3)
HIV diagnosed more than 6 weeks after	10 (6.8)	0
HIV diagnosis date not known	1	0
CD4 count (cells/mm^3)		
≤350	14 (16.7)	3 (20.0)
>350	70 (83.3)	12 (80.0)
Not known	65	9
Viral load (c/ml)		
≤50	74 (77.9)	13 (76.5)
>50	21 (22.1)	4 (23.5)
Not known	54	7

N=173 unless otherwise specified. Sexual identity only available for people with a questionnaire (n=55). Missing data excluded from percentage calculations, except for recent foreign travel. Recent foreign travel (past 4 days prior to symptoms) as recorded on questionnaire or on laboratory report forms for people who did not have a questionnaire.

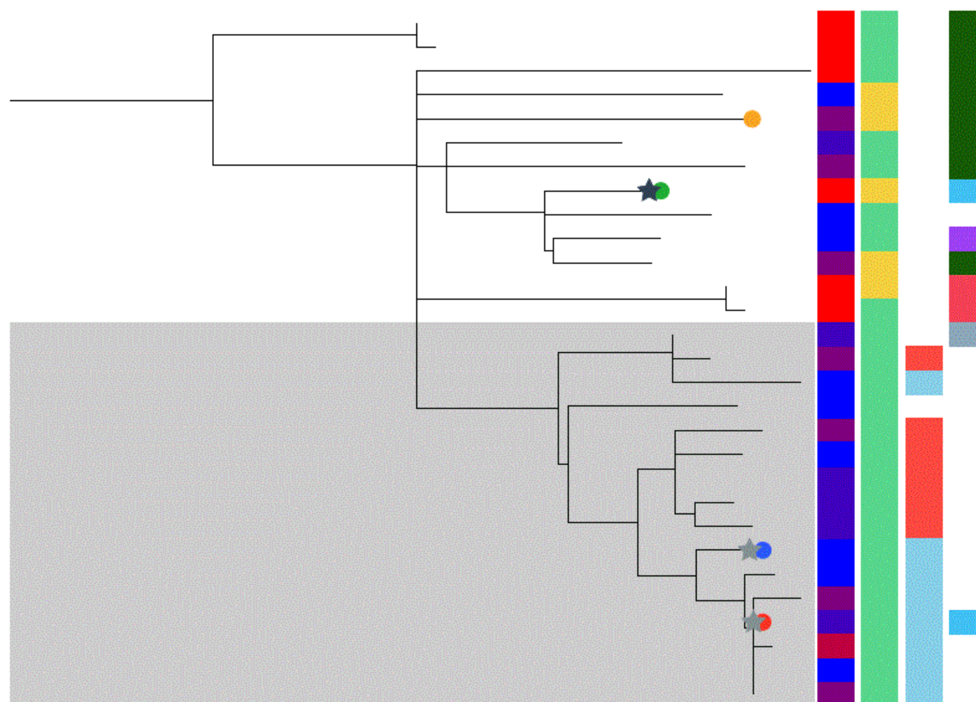
4.5.5 Detecting novel strain transmission in sexual networks of MSM

In addition to the two domestically circulating MSM clades discussed above, I used the combined WGS and epidemiological data to identify previously unknown lineages that, because of their epidemiological profile, were also likely being transmitted within sexual networks of MSM. I focused on isolates from confirmed MSM that were located phylogenetically within travel-associated lineages (Figure 4.3: confirmed MSM cases that did not fall into one of the two *S. flexneri* MSM clades and Table 4.5: details of these MSM cases). One exemplar cluster within PG2 was strongly suggestive of transmission in presumed MSM (highlighted in orange in Table 4.5). Phylogenetic analysis of all isolates within this cluster, contextualised using phylogenetically proximate isolates within 50 SNPs, revealed a previously unknown clade in presumed MSM (highlighted in grey, Figure 4.4). All isolates within this clade were from adult men, most of whom had not travelled abroad, and nearly all harboured genotypic markers of azithromycin resistance (87.5%; 14/16). Questionnaire data were available for two adult men in this clade, both of white ethnicity. One man self-identified as gay and one reported recent sexual contact but did not disclose their sexual identity or gender of their recent sex partner. By contrast, the other proximal isolates in the phylogenetic tree but not within this clade were from a mixed group of adult men, women and children, most of whom reported recent foreign travel (92.3%; 12/13). None of these isolates harboured genotypic markers of azithromycin resistance. Questionnaire data were available for two of the non-MSM clade cases including one adult female who had travelled to South Asia, and one child of Asian ethnicity who had recently travelled to the Middle East. I reconstructed a timed phylogeny which estimated that this novel lineage might have entered the MSM population approximately seven years ago (95% Highest Posterior Density 4 to 11 years) (Figure 4.5).

Table 4.5: Epidemiological and molecular characteristics of 10-SNP single linkage clusters± nested within travel-associated lineages and containing isolates from confirmed MSM

10-SNP single linkage cluster±	PG (serotype)	Confirmed MSM	Confirmed non-MSM	Total cases in cluster	M:F ratio	Foreign travel	*mphA* and/or *ermB*	Living with HIV**
78.324.644.966.1208.%	1 (3a)	1	0	2	2:0	0	1	1
3.47.85.237.397.%	**2 (1c)**	**1**	0	**10**	**10:0**	**2***	**10**	**3**
4.49.69.337.364.%	3 (2a)	2	0	2	2:0	1	0	0
4.49.69.307.330.%	3 (2a)	1	0	2	2:0	0	2	2
4.49.69.95.615.%	3 (2a)	1	0	1	1:0	0	0	1
4.49.49.281.297.%	3 (2a)	1	0	2	2:0	0	0	0
4.110.188.317.439.%	3 (2a)	1	0	2	2:0	0	2	1
4.110.162.273.287.%	3 (2a)	1	0	2	2:0	0	2	1
4.184.304.462.531.%	3 (2a)	1	0	1	1:0	0	0	0
42.115.171.292.308.%	3 (2a)	1	0	1	1:0	0	1	1

N=11 confirmed MSM cases as reported on questionnaire. ±A single representative from each 10-SNP single linkage cluster was presented in Figure 4.3. 10-SNP single linkage clusters with one case indicate that the isolate did not cluster with another isolate at the 10-SNP threshold. *Travel destination not recorded for 1 case. **Where reported, probable route of exposure to HIV reported as sex between men. Exemplar cluster showing a strong signal of an unknown lineage that is likely being transmitted in presumed MSM is highlighted in grey (3.47.85.237.397.%). PG: phylogenetic group.



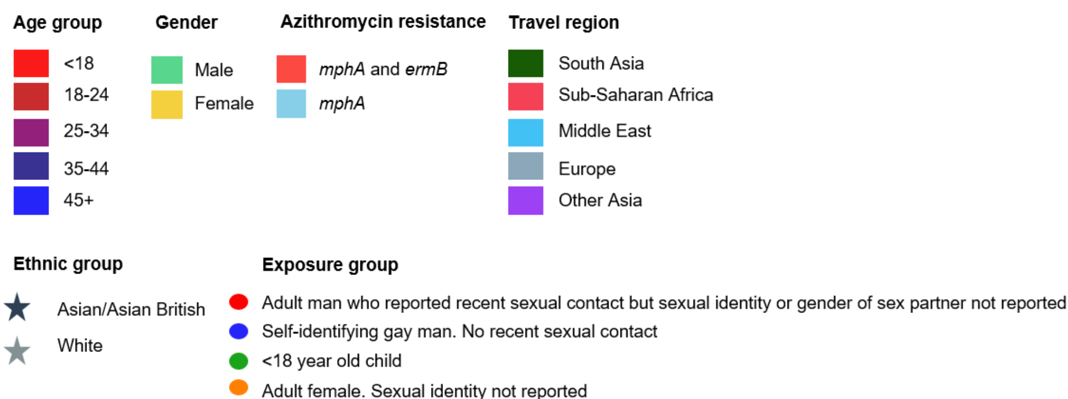


Figure 4.4: Detection of novel strain transmission among MSM

Mid-point rooted maximum likelihood phylogenetic tree of all isolates from one selected single linkage cluster at the 50-SNP threshold nested within a travel-associated lineage (PG2, serotype 1c, SNP address 3.47.85.%, n=29). Epidemiological data are represented as coloured strips (age group, gender, genotypic markers of azithromycin resistance) or as symbols on the branches (*S. flexneri* exposure group, ethnic group). Symbols are presented for people that have questionnaire data only. The clade associated with novel strain transmission in presumed MSM is highlighted in grey (n=16).

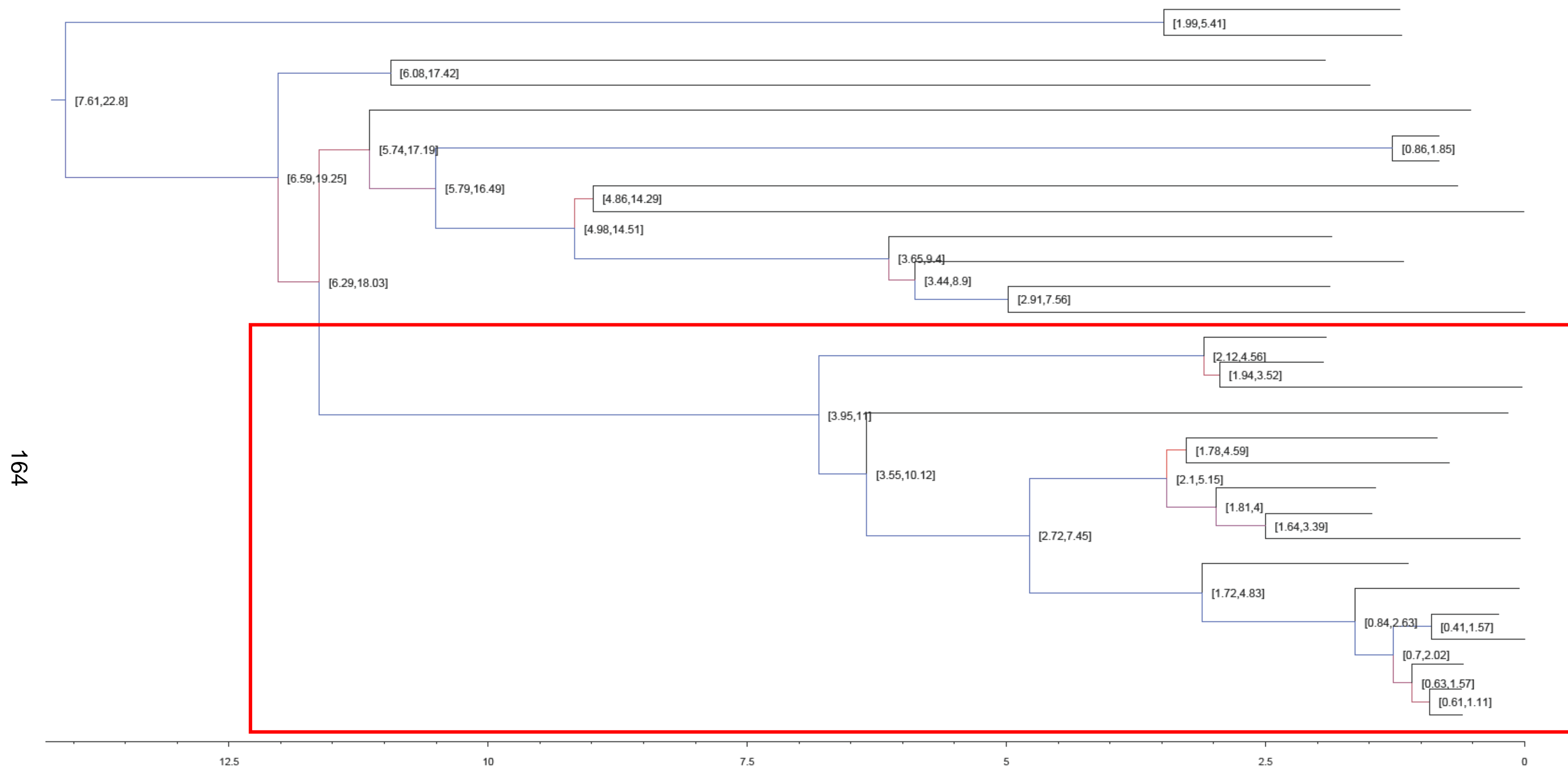


Figure 4.5: Timed phylogeny estimating the introduction of a novel lineage into a presumed MSM network

BEAST-generated maximum clade credibility tree of all isolates from one selected single linkage cluster at the 50-SNP threshold nested within a travel-associated lineage (PG2, serotype 1c, SNP address 3.47.85.%, n=29). The scale bar represents the time (years) since most recent common ancestor. Node ends show the 95% Highest Posterior Density estimates.

4.5.6 Clinical outcomes, treatment and health-seeking behaviour

4.5.6.1 Characteristics of cases with questionnaire data

Clinical characteristics, antimicrobial treatment and health-seeking behaviour among cases with questionnaire data are presented in Table 4.6. Where reported, 24.5% (40/163) of cases reported that they had previously heard of shigellosis and this was highest among MSM (36.0% [31/86]). Most cases sought healthcare from their GP or from a hospital service (e.g. A&E) and 21.7% (20/92) of MSM attended a SHC.

Where reported, 56.2% (95/169) had recovered from their illness (median duration 10 days [range 2 to 35 days]) and 43.8% (74/169) reported that they were still unwell (median duration at questionnaire completion 14 days [range 3 to 365 days]). Among those who had recovered, the median duration of symptoms was seven days (range 2-31) for confirmed MSM and 13 days (range 2-35) for confirmed other adults (p=0.003). Overall, 30.2% (55/182) of people reported being admitted to hospital. Six people were still hospital in-patients at the time of questionnaire completion (median duration of stay at questionnaire completion five days [range 3-15 days]). Among those who had been discharged and where information was available (n=41), the median duration of hospital stay was three days (range 1 to 10 days).

Information about antimicrobial class (questionnaire wording: “Treated with antibiotics” Responses: yes, no. If yes, specify) was poorly recorded. Among cases who reported that they had received antimicrobials, antimicrobial class was missing for 41.9% (54/129). Where recorded, most adults reported receiving ciprofloxacin alone (51.1% [22/43] for confirmed MSM, 76.2% [16/21] for confirmed other adults, 100% [3/3] for adults whose sexual identity and recent behaviour was not known). Nine confirmed MSM reported receiving more than one antimicrobial, of whom seven reported receiving ciprofloxacin in combination with at least one other antimicrobial and up to a maximum of three. Overall, six confirmed MSM were prescribed azithromycin (azithromycin only [n=3], azithromycin and doxycycline [n=2], azithromycin and ciprofloxacin [n=1]) and all had *S. flexneri* isolates harbouring genotypic markers of azithromycin resistance.

Table 4.6: Clinical characteristics, antimicrobial treatment and health-seeking behaviour among *S. flexneri* cases with questionnaire data

	MSM N=95	Other adults N=61	Children N=17	Not known* N=9
Previously heard of shigellosis±				
No	55 (64.0)	50 (87.7)	13 (92.9)	5 (83.3)
Yes	31 (36.0)	7 (12.3)	1 (7.1)	1 (16.7)
Not specified	9	4	3	3
Healthcare received				
GP	32 (34.8)	36 (62.1)	6 (37.5)	4 (57.1)
Hospital	28 (30.4)	9 (15.5)	3 (18.8)	2 (28.6)
Sexual Health/HIV clinic	13 (14.1)	1 (1.7)	0 (0)	1 (14.3)
GP & Hospital	12 (13.0)	12 (20.7)	7 (43.8)	0 (0)
GP, Hospital & sexual health/HIV clinic	7 (7.6)	0 (0)	0 (0)	0 (0)
Not specified	3	3	1	2
Diarrhoea				
No	1 (1.1)	1 (1.7)	0	0
Yes	92 (98.9)	59 (98.3)	16 (100)	9 (100)
Not specified	2	1	1	0
Abdominal pain				
No	6 (7.1)	6 (10.3)	4 (26.7)	1 (12.5)
Yes	78 (92.9)	52 (89.7)	11 (73.3)	7 (87.5)
Not specified	11	3	15	1
Vomiting				
No	54 (66.8)	36 (69.2)	6 (33.3)	4 (50.0)
Yes	27 (33.3)	16 (30.8)	10 (66.7)	4 (50.0)
Not specified	14	9	2	1
Fever				
No	17 (19.5)	20 (38.5)	3 (17.7)	2 (28.6)
Yes	80 (80.5)	32 (61.5)	14 (82.4)	5 (71.4)
Not specified	8	9	0	2
Mucus in stools				
No	31 (43.1)	34 (70.8)	5 (50.0)	3 (37.5)
Yes	41 (56.9)	14 (29.2)	5 (50.0)	5 (62.5)
Not specified	23	13	7	1
Blood in stools				
No	25 (29.4)	29 (52.7)	9 (69.2)	4 (50.0)
Yes	60 (70.6)	26 (47.3)	4 (30.8)	4 (50.0)
Not specified	10	6	4	1
Admitted to hospital				
No	55 (62.5)	46 (79.3)	9 (56.3)	6 (66.7)
Yes	33 (37.5)	12 (20.7)	7 (43.8)	3 (33.3)
Not specified	7	3	1	0
Antimicrobials prescribed				
No	18 (19.6)	25 (43.9)	0 (0)	1 (12.5)
Yes	74 (80.4)	32 (56.1)	16 (100.0)	7 (87.5)
Not specified	3	4	1	1

N=182. *Not known includes adult men who did not specify their sexual identity or provide information on recent sexual behaviour. ±Based on the question: Has the person heard of shigellosis/*shigella* spp. before? Missing data excluded from percentage calculations.

4.5.6.2 Associations with markers of clinical severity

Table 4.7 presents associations of epidemiological, clinical and molecular factors with the presentation of severe clinical symptoms (blood in stools, mucus in stools, fever) in adults diagnosed with *S. flexneri*. The age-adjusted odds of blood (adjusted odds ratio [aOR]: 2.18 [95% CI: 1.11 to 4.26]; p=0.022) or mucus (aOR: 2.49 [95% CI: 1.21 to 5.15]; p=0.011) in stools and fever (aOR: 2.58 [95% CI: 1.30 to 5.11]; p=0.006) were significantly higher for confirmed MSM compared to confirmed other adults.

There were also significant associations between clinical symptoms and explanatory factors that were highly correlated with being MSM or not. For instance, the age-adjusted odds of mucus in stools were significantly higher among people who were living with HIV at the time of their *S. flexneri* diagnosis compared to those who were HIV-negative/unknown status (aOR: 2.38 [94% CI: 1.21 to 4.70]; p=0.012), but there was no evidence for an association between HIV status and other clinical symptoms (blood in stools or fever). There were insufficient data available to explore whether there was an association between clinical symptoms and recent HIV viral load or CD4 cell count.

The age-adjusted odds of mucus in stools (aOR: 0.33 [95% CI: 0.16 to 0.67]; p=0.002) and fever (aOR: 0.45 [95% CI: 0.23 to 0.87]; p=0.017) were lower for people who had an isolate belonging to a travel-associated lineage compared to those who had an isolate belonging to one of the two domestically circulating MSM clades. Correspondingly, the age-adjusted odds of mucus in stools were significantly higher (aOR: 2.34 [95% CI: 1.18 to 4.61]; p=0.013) for people who had an isolate harbouring genotypic markers of azithromycin resistance. Finally, the age-adjusted odds of mucus in stools were lower for people of an ethnic minority group compared to those who were of white ethnicity (aOR: 0.35 [95% CI: 0.17 to 0.76]; p=0.006).

Table 4.8 presents associations of epidemiological, clinical and molecular factors with clinical outcomes (hospital admission, antimicrobial use) in adults diagnosed with *S. flexneri*. The age-adjusted odds of hospital admission (aOR: 2.20 [95% CI: 1.02 to 4.73]; p=0.038) and antimicrobial use (aOR: 3.27 [95% CI: 1.62 to 6.63]; p<0.001) were higher for confirmed MSM compared to confirmed other adults.

The association between clinical outcomes and explanatory factors that were correlated with being MSM or not was variable. For instance, the age-adjusted odds of antimicrobial use were higher for people who were living with HIV (aOR: 2.15 [95% CI: 1.00 to 4.64]; p=0.043) compared to people who were HIV-negative/unknown status, although this was of borderline significance, but there was no evidence for an association between HIV status and hospital admission. The age-adjusted odds of antimicrobial use were higher among people infected with isolates that harboured genotypic markers of azithromycin resistance compared to those who were infected with isolates without these markers (aOR: 2.82 [95% CI: 1.41 to 5.64]; p=0.003), and among people who had isolates that belonged to the two domestically circulating MSM clades (aOR for travel-associated lineage compared to MSM clades as the reference category: 0.46 [95% CI: 0.23 to 0.91]; p=0.025). By contrast, there was no evidence to suggest that azithromycin resistance or lineage were associated with hospital admission. However, infection with *S. flexneri* serotype 2a was strongly associated with hospital admission; the age-adjusted odds ratio for all other serotypes compared to serotype 2a as the reference category was 0.28 (95% CI: 0.11 to 0.72, p=0.004). The odds of hospital admission (aOR: 0.34 [95% CI: 0.15 to 0.75]; p=0.005) and antimicrobial use (aOR: 0.35 [95% CI: 0.18 to 0.69]; p=0.002) were also significantly lower for people who reported recent foreign travel. Furthermore, people of an ethnic minority group were less likely to report antimicrobial use compared to people of white ethnicity (aOR: 0.26 [95% CI: 0.12 to 0.54]; p<0.001).

Table 4.7: Characteristics associated with blood in stools, mucus in stools and fever in adults diagnosed with *S. flexneri* and with linked questionnaire data

	Blood in stools			Mucus in stools			Fever		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group (n=156)									
Other adults	26/61 (42.6)	1.00	1.00	14/61 (23.0)	1.00	1.00	32/61 (52.5)	1.00	1.00
MSM	60/95 (63.2)	2.31 (1.20-4.45)	2.18 (1.11-4.26)	41/95 (43.2)	2.55 (1.24-5.25)	2.49 (1.21-5.15)	70/95 (73.7)	2.54 (1.29-5.00)	2.58 (1.30-5.11)
p-value		0.012	0.022		0.009	0.011		0.007	0.006
HIV status (n=155)									
Negative/unknown	59/112 (52.7)	1.00	1.00	33/112 (29.5)	1.00	1.00	68/112 (60.7)	1.00	1.00
Living with HIV	30/52 (57.7)	1.22 (0.63-2.38)	1.19 (0.61-2.34)	26/52 (50.0)	2.39 (1.21-4.72)	2.38 (1.21-4.70)	38/52 (73.1)	1.76 (0.85-3.61)	1.76 (0.86-3.62)
p-value		0.548	0.608		0.012	0.012		0.119	0.118
Serotype (n=152)									
2a	60/105 (57.1)	1.00	1.00	38/105 (36.2)	1.00	1.00	72/105 (68.6)	1.00	1.00
Other	23/47 (48.9)	0.72 (0.36-1.43)	0.70 (0.35-1.43)	16/47 (34.0)	0.91 (0.44-1.87)	0.91 (0.44-1.88)	27/47 (57.5)	0.62 (0.30-1.26)	0.62 (0.30-1.26)
p-value		0.348	0.330		0.798	0.793		0.187	0.187
Lineage (n=165)									
MSM clade	58/97 (59.8)	1.00	1.00	45/97 (46.4)	1.00	1.00	70/97 (72.2)	1.00	1.00
Travel-associated lineage	32/68 (47.1)	0.60 (0.31-1.12)	0.67 (0.35-1.28)	15/68 (22.1)	0.33 (0.16-0.66)	0.33 (0.16-0.67)	37/68 (54.4)	0.46 (0.24-0.88)	0.45 (0.23-0.87)
p-value		0.106	0.227		0.001	0.002		0.019	0.017
Azr resistance (n=165)									
No	34/74 (46.0)	1.00	1.00	19/74 (25.7)	1.00	1.00	44/74 (59.5)	1.00	1.00
Yes	56/91 (61.5)	1.88 (1.01-3.51)	1.63 (0.86-3.10)	41/91 (45.1)	2.37 (1.22-4.62)	2.34 (1.18-4.61)	63/91 (69.2)	1.53 (0.81-2.92)	1.57 (0.81-3.03)
p-value		0.045	0.136		0.009	0.013		0.192	0.179

	Blood in stools			Mucus in stools			Fever		
Foreign travel (n=165)									
No	60/104 (57.7)	1.00	1.00	44/104 (42.3)	1.00	1.00	72/104 (69.2)	1.00	1.00
Yes	30/61 (49.2)	0.71 (0.38-1.34)	0.77 (0.40-1.47)	16/61 (25.2)	0.48 (0.24-0.97)	0.49 (0.25-0.99)	35/61 (57.4)	0.60 (0.31-1.15)	0.59 (0.31-1.15)
p-value		0.289	0.429		0.036	0.042		0.126	0.122
Age group (n=165)									
18-24	8/15 (53.3)	1.17 (0.39-3.48)	N/A	N/A	N/A	N/A	6/15 (40.0)	0.35 (0.11-1.07)	N/A
25-34	36/57 (63.2)	1.75 (0.89-3.44)		26/72 (36.1)§	0.98 (0.52-1.86)		40/57 (70.2)	1.23 (0.61-2.51)	
≥35	46/93 (49.5)	1.00		34/93 (36.6)	1.00		61/93 (65.6)	1.00	
p-value		0.258			0.953			0.101	
Per year (continuous variable)		0.97 (0.95-0.99)			0.99 (0.97-1.02)			1.00 (0.98-1.02)	
p-value (continuous variable)		0.008			0.448			0.977	
Ethnic group (n=147)									
White	59/97 (60.8)	1.00	1.00	46/97 (47.4)	1.00	1.00	66/97 (68.0)	1.00	1.00
Ethnic minority	24/50 (48.0)	0.59 (0.30-1.18)	0.61 (0.30-1.24)	12/50 (24.0)	0.35 (0.16-0.75)	0.35 (0.17-0.76)	31/50 (62.0)	0.77 (0.38-1.56)	0.77 (0.38-1.57)
p-value		0.138	0.171		0.005	0.006		0.466	0.472
IMD quintile (n=164)									
1-2 (most deprived)	70/125 (56.0)	1.00	1.00	47/125 (37.6)	1.00	1.00	84/125 (67.2)	1.00	1.00
3	10/23 (43.5)	0.60 (0.25-1.48)	0.70 (0.28-1.77)	9/23 (39.1)	1.07 (0.43-2.66)	1.11 (0.44-2.80)	18/23 (78.3)	1.76 (0.61-5.07)	1.72 (0.59-4.99)
4-5 (least deprived)	10/16 (62.5)	1.31 (0.45-3.83)	1.58 (0.53-4.77)	4/16 (25.0)	0.55 (0.17-1.81)	0.58 (0.17-1.91)	5/16 (31.3)	0.22 (0.07-0.68)	0.22 (0.07-0.67)
p-value		0.440	0.481		0.577	0.607		0.008	0.007

Total numbers vary for each question due to missing data. Azr: azithromycin resistance, IMD: Index of Multiple Deprivation. § Age categories for 18-24 and 25-34 years combined due to small numbers. Unadjusted and age-adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over.

Table 4.8: Characteristics associated with hospital admission and antimicrobial use in adults diagnosed with *S. flexneri* and with linked questionnaire data

	Hospital admission			Antimicrobial use		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group (n=156)						
Other adults	12/61 (19.7)	1.00	1.00	32/61 (52.5)	1.00	1.00
MSM	33/95 (34.7)	2.17 (1.02-4.65)	2.20 (1.02-4.73)	74/95 (77.9)	3.19 (1.59-6.42)	3.27 (1.62-6.63)
p-value		0.039	0.038		0.001	<0.001
HIV status (n=164)						
Negative/unknown	30/112 (26.8)	1,00	1.00	71/112 (63.4)	1.00	1.00
Living with HIV	17/52 (32.7)	1.33 (0.65-2.71)	1.33 (0.65-2.71)	41/52 (78.9)	2.15 (1.00-4.64)	2.15 (1.00-4.65)
p-value		0.439	0.443		0.043	0.043
Serotype (n=152)						
2a	36/105 (35.0)	1.00	1.00	72/105 (68.6)	1.00	1.00
Other	6/47 (12.8)	0.28 (0.11-0.72)	0.28 (0.11-0.72)	30/47 (63.8)	0.81 (0.39-1.67)	0.81 (0.39-1.67)
p-value		0.004	0.004		0.567	0.567
Lineage (n=165)						
MSM clade	31/97 (32.0)	1.00	1.00	73/97 (75.3)	1.00	1.00
Travel-associated lineage	17/68 (25.0)	0.71 (0.35-1.42)	0.71 (0.35-1.45)	40/68 (58.8)	0.47 (0.24-0.92)	0.46 (0.23-0.91)
p-value		0.330	0.346		0.026	0.025
Azithromycin resistance (n=165)						
No	21/74 (28.4)	1.00	1.00	42/74 (56.8)	1.00	1.00
Yes	27/91 (29.7)	1.06 (0.54-2.09)	1.05 (0.52-2.09)	71/91 (78.0)	2.70 (1.38-5.32)	2.82 (1.41-5.64)
p-value		0.856	0.897		0.003	0.003

	Hospital admission			Antimicrobial use		
Foreign travel (n=165)						
No/unknown	38/104 (36.5)	1.00	1.00	80/104 (76.9)	1.00	1.00
Yes	10/61 (16.4)	0.34 (0.16-0.75)	0.34 (0.15-0.75)	33/61 (54.1)	0.35 (0.18-0.70)	0.35 (0.18-0.69)
p-value		0.005	0.005		0.003	0.002
Age group (n=165)						
18-24	3/15 (20.0)	0.68 (0.18-2.61)	0.68 (0.18-2.61)	10/15 (66.7)	0.86 (0.27-2.75)	0.86 (0.27-2.75)
25-34	20/57 (35.1)	1.47 (0.72-3.00)	1.47 (0.72-3.00)	38/57 (66.7)	0.86 (0.42-1.75)	0.86 (0.42-1.75)
≥35	25/93 (26.9)	1.00	1.00	65/93 (69.9)	1.00	1.00
p-value		0.401	0.401		0.907	0.907
Per year (age as a continuous variable)		1.00 (0.97-1.02)	1.00 (0.97-1.02)		1.00 (0.97-1.02)	1.00 (0.97-1.02)
p-value (age as a continuous variable)		0.793	0.793		0.930	0.930
Ethnic group (n=147)						
White	26/97 (26.8)	1.00	1.00	77/97 (79.4)	1.00	1.00
Ethnic minority	17/50 (34.0)	1.41 (0.67-2.94)	1.42 (0.68-2.97)	25/50 (50.0)	0.26 (0.12-0.55)	0.26 (0.12-0.54)
p-value		0.367	0.358		<0.001	<0.001
IMD quintile (n=164)						
1-2 (Most deprived)	37/125 (29.6)	1.00	1.00	85/125 (68.0)	1.00	1.00
3	8/23 (34.8)	1.27 (0.50-3.24)	1.28 (0.50-3.32)	17/23 (73.9)	1.33 (0.49-3.64)	1.34 (0.48-3.68)
4-5 (Least deprived)	3/16 (18.8)	0.55 (1.15-2.04)	0.56 (0.15-2.08)	11/16 (68.8)	1.04 (0.34-3.18)	1.04 (0.33-3.21)
p-value		0.531	0.534		0.850	0.850

Total numbers vary for each question due to missing data. IMD: Index of Multiple Deprivation. Unadjusted and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over.

I also explored the characteristics associated with two composite outcomes of severity: clinical symptoms (blood and/or mucus in stools) and clinical outcomes (hospital admission and/or antimicrobial use) (Appendix 4.4). After adjusting for age, there was weak evidence to suggest that confirmed MSM were more likely to present with clinical symptoms compared to non-MSM (aOR: 1.88 [95% CI: 0.95 to 3.73]; p=0.070). The age-adjusted odds of clinical symptoms were significantly lower for people of an ethnic minority group compared to those of white ethnicity (aOR: 0.46 [95% CI: 0.22 to 0.95]; p=0.036).

The age-adjusted odds of clinical outcomes were significantly higher among MSM compared to non-MSM (aOR: 3.07 [95% CI: 1.49 to 6.31]; p=0.002). Additionally, the age-adjusted odds of clinical outcomes were significantly higher among people living with HIV compared to people who were HIV-negative/unknown status (aOR: 2.50 [95% CI: 1.08 to 5.57]; p=0.024) and in people who had isolates harbouring genotypic markers of azithromycin resistance (aOR: 2.75 [95% CI: 1.35 to 5.61]; p=0.005). Consistent with the above, clinical outcomes were less likely to be reported by people of an ethnic minority group compared to those of white ethnicity (aOR: 0.39 [95% CI: 0.18 to 0.82]; p=0.013), by those who reported recent foreign travel compared to those that did not (aOR: 0.34 [95% CI: 0.17 to 0.68]; p=0.002), and in people who had an isolate belonging to a travel-associated lineage compared to an MSM clade (aOR: 0.49 [95% CI: 0.25 to 0.99]; p=0.045), although the latter was of borderline significance (Appendix 4.4).

4.5.6.2.1 Sensitivity analyses excluding missing data

Sensitivity analyses excluding people with missing data on clinical symptoms and outcomes did not substantially change the results. The measures of association tended to be stronger, but confidence intervals were wider (Appendices 4.5 to 4.7).

4.6 Discussion

4.6.1 Summary of key findings

Using phylogenetic analyses, overlaid with behavioural and clinical data, I found that 43.0% (398/926) of CC245 *S. flexneri* cases in England during August 2015 to July 2017 belonged to two domestically circulating MSM clades that were associated with the presence of genotypic markers of azithromycin resistance. 88.4% of isolates taken from confirmed MSM belonged to these clades, indicating that clonal expansion most likely occurred through sexual transmission. Furthermore, over one third (135/398) of isolates within these MSM clades were from people living with HIV, indicating the overlap between these epidemics. In addition to the two MSM clades, my analysis revealed strains belonging to multiple other genetic lineages of *S. flexneri* that are being transmitted within sexual networks of presumed MSM.

One in five MSM belonged to a recognised group that were at increased risk of spreading their infection to other people (see section 4.4.2). Despite this, there was no evidence of sustained transmission beyond sexual networks. Most cases sought healthcare from their GP or a hospital service (e.g. A&E), and only a small proportion of MSM attended a SHC (one in five). Overall awareness of shigellosis in the study population was low, but was higher among confirmed MSM compared to other adults (36.0% vs 12.3%). Confirmed MSM were more likely to report severe clinical symptoms and/or outcomes compared to other adults, but the exact reasons for this remain unclear. Ciprofloxacin was the main antimicrobial prescribed among the study population, but some confirmed MSM received azithromycin, even though their isolates were resistant to this antimicrobial.

4.6.2 Strengths and limitations

The study builds on and improves upon previous molecular epidemiology studies on cases of domestically acquired shigellosis in England, which have until now included very limited data on sexual identity and behaviour, as well as clinical factors such as HIV co-infection. One of the key strengths of this study was that I included all *S. flexneri* isolates from England over a two-year period that were submitted to the national reference laboratory, which improves the

generalisability of my findings. Another strength was that I had rich epidemiological data for nearly one in five isolates that was collected through an exposure questionnaire compared to previous studies that used basic demographic data from laboratory report forms (age, gender and recent foreign travel history). For the first time in England, the questionnaire included information on sexual identity and behaviour. By combining comprehensive WGS data with socio-demographic, clinical and behavioural data, I have provided unique insights into the distribution and characteristics associated with different *S. flexneri* transmission networks in England. Additionally, this is the first study to describe clinical outcomes, healthcare seeking behaviour and awareness of infection for a large sub-set of cases in England and how these differ between MSM and other adults.

The main limitation for this study was that questionnaire data were only available for *S. flexneri* cases reported to the HPTs who participated in the pilot of the new standardised questionnaire (see section 4.4.2). Compared to all isolates included in the study without questionnaire data, isolates with linked questionnaire data were more likely to come from cases living in London. In addition, a higher proportion of isolates with linked questionnaire data were taken from adult cases compared to other cases, or from those who had not travelled abroad compared to those who did travel. It is likely that this reflects the demographic population of the participating HPTs, including London, Brighton, Manchester, as these regions have proportionally higher MSM populations where *S. flexneri* is likely acquired domestically through sex between men.[298] In addition, clinical symptoms and outcomes were based on self-reporting. It is possible that MSM perceived or reported their symptoms differently to non-MSM, particularly if these were related to other co-infections or co-morbidities, such as those associated with STIs and HIV.

My study was based on surveillance data of *S. flexneri* cases who presented to healthcare settings and had a stool sample collected for investigation by local laboratories that was referred to the GBRU for molecular typing. It is estimated that approximately two thirds of specimens with *Shigella* spp. isolated at local hospital laboratories are referred to the GBRU but the number of undiagnosed cases in the wider community is unknown and likely to be large.[200] This under-ascertainment might have led to bias in my study because many factors

are likely to influence whether an isolate was included. The IID1 study in England found that presenting to healthcare settings was positively associated with severity of illness, recent foreign travel, leaving full-time education at a younger age and lower socioeconomic status,[254] and consequently these characteristics might be over-represented in my study.

4.6.3 Interpretation of the results

4.6.3.1 Molecular epidemiology of S. flexneri

Among cases with questionnaire data, nearly half were confirmed MSM, and I observed strikingly different epidemiological and molecular characteristics for confirmed MSM compared to confirmed non-MSM. The main exposure for confirmed MSM was recent sex with a man (72.3% of MSM compared to 50.0% of heterosexual men who had recent sex with a woman), while the main exposure for confirmed non-MSM was foreign travel (76.9% of non-MSM cases had visited regions traditionally considered to be at high-risk for shigellosis e.g. South Asia and sub-Saharan Africa, compared to 2.1% of confirmed MSM). Despite these prevailing trends, I also observed that 27.7% of confirmed MSM men did not report recent sexual contact, which might be due to transmission through non-sexual contact, a delay in symptom onset due to an incubation period of longer than four days, relapse of a previously acquired chronic infection, or failure to report recent behaviour.

Although I observed *S. flexneri* isolates from confirmed MSM to be distributed throughout the phylogeny, 88.4% (84/95) of isolates belonged to two major domestically circulating clades that were described by Baker *et al.* to be MSM-associated based on circumstantial demographic data (*S. flexneri* serotype 3a [2015] and *S. flexneri* serotype 2a [2018]).[22,168] The high proportion of confirmed MSM isolates in my study belonging to these clades is consistent with rapid and sustained transmission within a large sexual network of MSM. In addition, I found that over one third (37.4%) of isolates within these clades belonged to people who were living with diagnosed-HIV, primarily acquired through sex between men, which suggests that *S. flexneri* infection likely occurs in overlapping sexual networks where people may be at risk of acquiring other STIs and HIV.

I found that MSM clades were strongly associated with *S. flexneri* isolates that harboured genotypic markers of azithromycin resistance, which may play a role in propagating epidemic expansion.[24] Using nationally representative sub-sets of isolates, Baker *et al.* (2015, 2018) found dramatic increases in case numbers (PG1, serotype 3a in 2010 and PG3, serotype 2a in 2012) following horizontal transfer of a specific plasmid carrying *mphA* and *ermB* (pKSR100).[24] My analysis shows that these *S. flexneri* MSM clades have continued to expand and were associated with 43.0% of all *S. flexneri* CC245 cases occurring in England during a two-year period from August 2015 to July 2017. These clades have also been described in MSM populations elsewhere,[168,169] highlighting the importance of international travel in connecting sexual networks and facilitating global dissemination. Rapid and sustained transmission of *Shigella* spp. is not typically observed in non-MSM populations in the UK, among whom exposure is primarily through travel and onward person-to-person transmission is unlikely to persist over long periods of time. However, there are some similarities between the spread of *S. flexneri* in MSM with outbreaks occurring within dense populations of the Orthodox Jewish Community (OJC) through person-to-person transmission in nursery and household settings.[299] Global outbreaks of *S. sonnei* in the OJC community, although not linked to sexual activity, have been associated with a single monophyletic lineage which emerged in Israel in the 1980s and spread across Europe and North America.[300]

I found that there were some non-MSM cases within the *S. flexneri* MSM clades, including women (1.5% of all isolates within the MSM clades) and children (1.0% of all isolates within the MSM clades), but it was not possible to elucidate direct transmission, or the direction of transmission. Where questionnaire data were available, I observed six heterosexual men and seven other adult men who did not provide information on identity or behaviour and these men had isolates phylogenetically located within the MSM clades. These men may represent MSM who have not disclosed sex with men, including heterosexual-identifying MSM, possibly due to the sensitive nature of the questions and perceived stigma,[301,302] or men who have acquired *S. flexneri* through non-sexual transmission. Together, these data suggest that there is some overlap between MSM and non-MSM transmission but do not provide evidence of significant or sustained transmission beyond sexual networks.

The combination of questionnaire and phylogenetic data allowed me to identify unknown lineages within presumed MSM networks which were nested within travel-associated lineages. These phylogenetically distinct lineages are likely to have been introduced to the MSM population through travel to a high-risk region or non-sexual transmission, followed by onward transmission through sexual contact. For one lineage in my dataset, there were pronounced differences in the patterns of gender, travel history, and genotypic markers of AMR, as well as differences in ethnicity and sexual identity based on the questionnaire data, which together were consistent with sexual transmission in presumed MSM for one clade and travel-associated transmission in phylogenetically proximal isolates. I estimated that this lineage was introduced into a presumed MSM sexual network approximately seven years ago. Given the rise in case numbers associated with MSM clades that harbour genotypic markers of azithromycin resistance, this lineage is one that should be monitored because it could potentially expand and become endemic.

4.6.3.2 Clinical severity of S. flexneri infection

In my study, I observed that confirmed MSM were more likely than confirmed other adults to report blood or mucus in stools, fever, admission to hospital or antimicrobial use. This is consistent with findings from *Shigella* spp. surveillance data from the USA which suggested higher odds of severe *S. flexneri* (defined as hospitalisation, bacteraemia or death associated with *S. flexneri* infection of any serotype) in adult men compared to women, although sexual identity and behaviour were not assessed.[303] I also found that there were strong positive associations between clinical symptoms/outcomes and explanatory variables that were highly correlated with being an MSM including HIV coinfection, serotype (i.e. MSM were more likely to be infected with *S. flexneri* isolates that were of serotype 2a compared to other adults), the molecular characteristics of the infecting lineage (e.g. MSM were more likely to have isolates that belonged to an *S. flexneri* MSM clade, or to have isolates that harboured genotypic markers of azithromycin resistance) and epidemiological characteristics (e.g. MSM were less likely to report recent foreign travel and were more likely to be of white ethnicity compared to other adults). Due to the smaller size sample and the high level of correlation between explanatory variables, I did not conduct a multivariable analysis to explore the factors associated with clinical severity controlling for potential confounding. When explanatory

variables are highly correlated, including them together in a model can result in large changes to the coefficient estimates, lower precision and lack of statistical significance. Such multicollinearity presents a significant challenge to interpreting regression models as it is no longer possible to distinguish the actual effect of each included factor.

In my study, HIV status was associated with the presence of mucus in stools, antimicrobial use, and with a composite measure of clinical severity (antimicrobial use and/or hospital admission), but there were insufficient data to enable me to explore the association with CD4 count and HIV viral load. Where data were available however, most had an undetectable viral load (≤50) and a high CD4 count (>350 cells/mm^3). HIV has been described as a risk factor for shigellosis[120-122] and several case studies found that HIV-related immunosuppression was associated with more severe illness, although most of these were reported prior to the introduction of highly-active antiretroviral therapy.[127-129] HIV causes substantial damage to the gut mucosal barrier at an early stage of infection due to the preferential depletion of CD4 cells in the gut-associated lymphoid tissue (GALT) accompanied by increased levels of immune activation.[304] Despite suppressive antiretroviral therapy, evidence suggests that complete restoration of mucosal immunity does not always occur, although the clinical consequences of this are unclear.[304,305] On the other hand, the association with HIV could reflect disparities in symptom reporting by HIV status, particularly if people living with HIV are more likely to worry about symptoms and/or co-infections, or to misreport symptoms that might not be related to their *S. flexneri* infection (e.g. mucus in stools that might be caused by LGV). Additionally, people with HIV may be more likely to attend healthcare and have stool specimens collected for microbiological investigations, which could have accounted for some of association observed in this population. Clinicians might also be inclined to prescribe antimicrobials for shigellosis if the case is living with HIV to avoid further complications.

I found that isolates belonging to MSM clades, as well as isolates harbouring genotypic markers of azithromycin resistance were associated with markers of clinical severity including mucus in stools, antimicrobial usage and fever. Increased clinical severity has been associated with phylogenetic lineage among other BEPs. In England, for example, the clinical severity (defined as bloody diarrhoea, hospital admission, death or HUS) of Shiga toxin-producing *E.*

coli (STEC) infection was strongly associated with sub-lineage after adjusting for age group and sex.[306] Furthermore, I found that hospital admission was strongly associated with *S. flexneri* serotype 2a compared to other *S. flexneri* serotypes, which could reflect the presence of specific virulence determinants that have previously been associated with serotype 2a.[47,297]

Other factors that might increase the likelihood of severity in MSM include the infecting dose. For instance, a meta-analysis of salmonellosis outbreaks exploring the association between proxy measures of infectious dose (amount of food, type of vehicle, attack rate and incubation period) and hospitalisation rates found evidence for a dose-severity relationship.[307] In my study, it is biologically plausible that direct oral-anal contact results in a large infectious dose when compared to other transmission routes such as indirect faecal-oral transmission through exposure to contaminated surfaces. Additionally, recent studies have shown that the composition of the gut microbiome can differ between MSM and non-MSM, regardless of HIV status,[308-310] which could result in increased rectal or systemic inflammation, but studies have yielded inconsistent results.[308,309,311]

4.6.4 Public health implications

My study shows that WGS data combined with epidemiological data provides unique insights into the likely route of *S. flexneri* infection (i.e. sexual or non-sexual), which could be used to inform contact tracing and targeted prevention. Timely understanding of the likely transmission route is important to differentiate between faecal-oral transmission, which might require a rapid environmental health response (e.g. closing a restaurant or isolating a case who may be leading to secondary infections), and sexual transmission, where patients might benefit from referral to sexual health services for further STI or HIV testing, partner notification and appropriate clinical management.

An important finding in my study was that sexual identity and sexual behaviour were well completed in the questionnaire and there was a high level of consistency between the two variables. The high level of completion with regards to sensitive questions relating to sexuality and recent sexual behaviour demonstrates that people are willing to answer these questions in the context of public health follow-up and that they should be included in surveillance

questionnaires to inform infection control and public health management. That said, some people, such as heterosexual-identifying MSM, may still not disclose sensitive and potentially stigmatising sexual behaviours and may not engage with health prevention messages targeting gay-identifying MSM.[312,313] Extending data collection for recent sexual contact to all adult cases (rather than just men) and to cover a longer period of time (e.g. the past seven days rather than four days) would provide a more detailed exposure history for cases that could improve infection management protocols, e.g. by identifying potential sexual exposure in MSM outside the 'accepted' four-day exposure period and in women whose isolates were phylogenetically located within MSM clades.

Most MSM sought care from their GP or hospital service (e.g. A&E), while only around one fifth reported accessing care through a SHC. The extent to which clinicians routinely consider sexual transmission as a route of infection within non-specialist settings is unknown, and many likely assume other exposure routes, as was reported during an investigation into a national outbreak of *S. flexneri* serotype 3a in 2012-2013 in the UK.[4] MSM may also not be aware of the potential for sexual transmission of their infection, and may not consider their sexual identity or behaviour relevant to their illness. I found that awareness of shigellosis amongst MSM in my study was only 36.0%. Improving awareness in both patients and health-care professionals is essential to control transmission.

Previous studies have found that MSM diagnosed with shigellosis are often co-infected with or have a recent history of having other bacterial STIs.[4] Consequently, off-target exposure to antimicrobials administered for other bacterial STIs has likely resulted in the development of azithromycin resistance in MSM-associated strains of *Shigella* spp., since azithromycin is not the primary treatment for shigellosis.[22] My study showed that 83.2% of confirmed MSM were infected with isolates that harboured genotypic markers of azithromycin resistance. Although antimicrobial usage was poorly completed in the questionnaire, some MSM infected with azithromycin-resistant *S. flexneri* isolates were prescribed azithromycin (14.0%, 6/43), either alone or in combination with other antimicrobials. Unfortunately, the questionnaire did not specify the indications for treatment and azithromycin may have been prescribed for a co-infection (e.g. bacterial STI). Additionally, nine MSM in my study reported receiving more than

one antimicrobial, which might suggest they had multiple infections. Despite the limitations of the antimicrobial usage data, however, better antimicrobial stewardship in MSM and the importance of taking a holistic approach that considers the long-term consequences of frequent antimicrobial exposure in high-risk populations, specifically the development of resistance in both target and non-target pathogens, is warranted. The collection of data on antimicrobial usage should be improved in the future given the increasing concerns about AMR in MSM and the wider population. To facilitate this, HPTs or EHOs may want to speak directly to the attending clinicians to confirm the specific course of antimicrobials taken. In addition, the questionnaire wording could be changed so that it is more precise.

4.6.5 Chapter summary

This chapter aimed to identify and characterise *S. flexneri* transmission in sexual networks of MSM and to explore how sexual transmission overlaps with non-sexual transmission within the community using linked WGS and epidemiological data. In summary:

- WGS data combined with rich epidemiological data has revealed the introduction, dissemination and persistence of strains within MSM sexual networks and how this overlaps with non-sexual transmission networks. *S. flexneri* isolates taken from confirmed MSM were distributed throughout the phylogeny but were primarily associated with two domestically circulating MSM clades (PG3, serotype 2a and PG1 serotype 3a), indicating the narrow phylogenetic diversity of most *S. flexneri* isolates circulating among MSM in England. There was some overlap between confirmed MSM and non-MSM cases but no evidence of sustained transmission beyond sexual networks.
- A high proportion of *S. flexneri* isolates within MSM clades harboured genotypic markers of azithromycin resistance. Some confirmed MSM were prescribed azithromycin for their *S. flexneri* infection, even though their isolates were resistant to this antimicrobial.
- Nearly 20% of *S. flexneri* isolates were taken from people living with HIV, of which 86.1% fell phylogenetically within the MSM clades, revealing the overlap between these infections. HIV preceded the shigellosis diagnosis in most cases.

- Compared to other adults, confirmed MSM were more likely to report blood or mucus in stools, fever, admission to hospital and antimicrobial use, indicating that their *S. flexneri* infections were potentially more severe. The exact reasons for this are unclear and require further exploration.

Chapter 5: Validation of a SNP clustering algorithm for identifying sexual transmission of *Shigella flexneri* in MSM

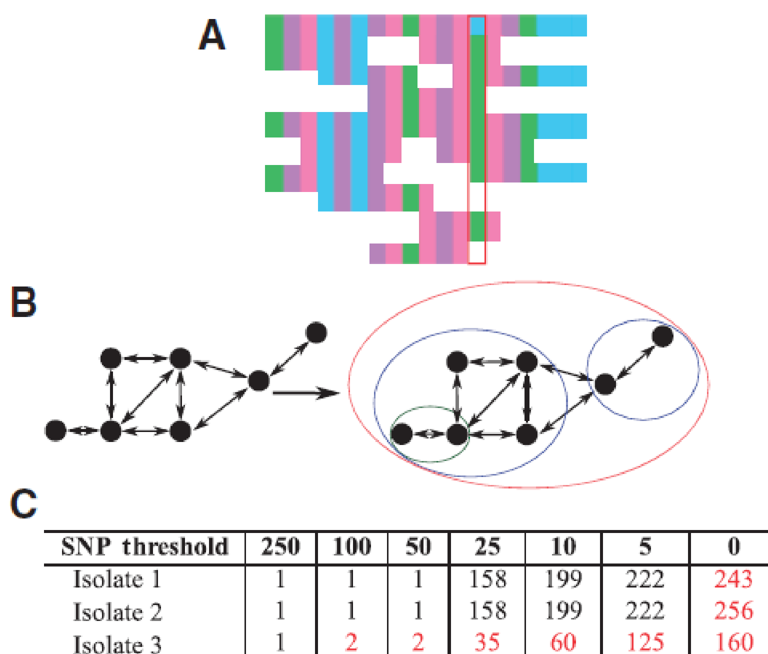
One of the major challenges that comes with the generation of whole genome sequencing (WGS) data for public health decision-making purposes is the large volume of data, which requires rapid synthesis, analysis and translation into epidemiologically useful information. To address this challenge, PHE has developed a Single Nucleotide Polymorphism (SNP) clustering algorithm to facilitate real-time cluster detection. In this chapter, I use the linked WGS and epidemiological dataset described in Chapter 4 to determine whether the SNP clustering algorithm can rapidly and accurately identify clusters of cases associated with sexual transmission in MSM and thereby inform appropriate infection control responses. The findings from this chapter have been published in a peer-reviewed open-access scientific journal distributed under the terms of the Creative Commons Attribution License (CC-BY) and subject to Crown copyright: Mitchell HD, Mikhail AFW, Painset A, Dallman TJ, Jenkins C, Thomson NR, Field N, Hughes, G. Use of whole genome sequencing to identify clusters of *Shigella flexneri* associated with sexual transmission in men who have sex with men in England: a validation study using linked behavioural data. *Microbial Genomics* 2019;5(11).

5.1 Introduction and rationale

Since August 2015, WGS has been performed on all isolates of *Shigella* spp. referred to PHE's national reference laboratory (the Gastrointestinal Bacteria Reference Unit [GBRU]).[45,269] SNP typing is used to aggregate isolates into clusters considered to represent cases that are linked through recent transmission events. The SNP typing tool informs public health decision-making in near real-time by distinguishing linked cases that might be part of an evolving UK outbreak and requiring a rapid and robust infection control response, from isolated cases such as those where infection was likely acquired during travel abroad.[273]

Hierarchical single linkage clustering is performed on the pairwise SNP distance matrix at descending SNP thresholds (250, 100, 50, 25, 10, 5, 0).[273] Isolates are included in a cluster if their pairwise genetic distance is within the SNP threshold to at least one other isolate (i.e.

single linkage). Thus, the maximum SNP difference between two isolates in any given cluster may be greater than the threshold. Clustering is summarised as a ‘SNP address’ (a seven-digit code) which describes the cluster membership at each of the thresholds (Figure 5.1).[45,273] If two isolates have a matching ‘SNP address’, there are no SNP differences and the isolates are said to be genetically indistinguishable.[45] Matching SNP addresses provide the strongest level of genetic evidence that isolates are likely linked by a common exposure or by direct person-to-person transmission. A threshold of 10-SNPs difference across the core genome sequences of any two isolates is the current standard for defining likely transmission clusters for routine public health surveillance of *Shigella* spp. at PHE.



SNP threshold	250	100	50	25	10	5	0
Isolate 1	1	1	1	158	199	222	243
Isolate 2	1	1	1	158	199	222	256
Isolate 3	1	2	2	35	60	125	160

Figure 5.1: Illustration of the SNP address

A. Illustration of a SNP difference between a reference sequence (top row) and a set of isolate sequences (remaining rows). B. Single linkage clustering of SNP differences into 0 (green, innermost circle), 5 (blue, two midmost circles) and 10 (red, outermost circle) SNP thresholds for a set of isolates. C. Examples of SNP addresses based on the seven descending SNP thresholds (250, 100, 50, 25, 10, 5, 0). Red digits indicate differences in the cluster membership at each SNP threshold. Source: Dallman *et al.* (2018).[273] Contains public sector information licensed under the Open Government Licence v3.0.

Limited routine demographic data are submitted to the GBRU alongside laboratory isolates (sex, date of birth, postcode of residence and foreign travel history) and used to classify clusters and infer the likely route of infection. Clusters are designated as ‘adult male’ if they

comprise (1) between two and five cases in total of which all are men aged 16 years or older, or (2) more than five cases where at least 90% are men aged 16 years or older. Clusters are designated as 'household' if two or more cases share a living space (i.e. either same postcode of residence or have the same surname and are identified in the same HPT area if the residential postcode is unavailable). Clusters are designated 'travel-associated' if they contain two or more cases and at least 50% report travel to the same country (or world region if country is not reported) outside the UK. 'Community' clusters are those that do not meet any of the other classifications, including clusters of between two and five cases where at least one case is a woman, and clusters of more than five cases where the proportion of men aged 16 years or older is less than 90%. Where data on sexual identity and behaviour are not routinely or rapidly available, 'adult male' clusters are used as a pragmatic proxy for sexual transmission in MSM. However, this approach has never been validated. In this chapter I aimed to address the following research question:

- How accurate is real-time SNP typing for identifying emergent clusters of *Shigella* spp. associated with sexual transmission in MSM?

5.2 Aim and Objectives

In this chapter I aimed to validate whether real-time WGS SNP typing can distinguish clusters representing sexual transmission in MSM from clusters representing non-sexual transmission to facilitate rapid and appropriate infection control responses. The objectives were to:

1. Determine whether clusters classified as 'adult male' represent likely sexual transmission in MSM by using information not previously available on sexual identity and sexual behaviour from linked case questionnaires.
2. Explore the sensitivity and specificity of the SNP clustering algorithm at a range of SNP thresholds for distinguishing clusters representing sexual transmission in MSM from non-sexual transmission clusters.
3. Explore whether genotypic markers of azithromycin resistance are a marker of clusters associated with sexual transmission in MSM.

5.3 Methods

5.3.1 Data sources

The combined WGS and epidemiological dataset used for this analysis was described in Chapter 4. In brief, I used WGS data of all *S. flexneri* isolates referred to the PHE reference laboratory between August 2015 and July 2017. Demographic data submitted alongside laboratory isolates were extracted from GDW (an isolate-level database storing GBRU laboratory results), and information on sexual identity and behaviour were available for a sub-set of isolates, collected through a pilot of a new standardised questionnaire used as part of routine follow-up for shigellosis (August 2015 to March 2017).

5.3.2 Cluster classification

To classify clusters, I used a computer programming script previously developed by the Gastrointestinal Infections Surveillance Department at PHE. Clusters included two or more isolates whose pairwise SNP distance was within a defined threshold and summarised as the 'SNP address'. I performed the cluster classification at a range of SNP thresholds (0, 5, 10 and 25). Clusters were classified as 'adult male', 'household', 'travel-associated' or 'community' using routine demographic data submitted alongside laboratory cultures, as described in section 5.1.

5.3.3 Describing national cluster data

To describe the cluster data available nationally during the study period, I used the current standard threshold of 10-SNPs difference across the core genome sequences of any two isolates. Differences in cluster size and duration (i.e. time between the first and last reported cases) between 'adult male' clusters and other non-sexual transmission clusters were assessed using the Chi-squared test for comparing two proportions.

5.3.4 SNP validation analysis

For all clusters that contained at least one case with a completed questionnaire, I assessed overall sensitivity and specificity of the cluster tool for each SNP threshold (0, 5, 10 and 25) using the questionnaire data as the gold standard to classify cases as either MSM (gay or bisexual-identifying men, or men who reported recent same-sex sexual contact) or non-MSM (heterosexual men, women and children under the age of 18 years); men who did not report any information on sexual identity and recent sexual contact were excluded. Sensitivity was defined as the proportion of MSM that belonged to an 'adult male' cluster and specificity was defined as the proportion of non-MSM that belonged to a non-sexual transmission cluster ('community', 'travel-associated' or 'household'). I calculated the proportion of 'adult male' clustered cases that (1) self-identified as gay or bisexual men, (2) reported recent same-sex sexual contact, and (3) reported recent foreign travel at a range of SNP thresholds. I then explored the impact of the SNP threshold on the classification, size and distribution of the cluster.

5.3.5 Antimicrobial resistance

For all 10-SNP clusters that contained at least one individual with a completed questionnaire, I explored the proportion of cases with WGS data that showed genotypic markers of azithromycin resistance (*mphA* and *ermB*). The questionnaire data were used to calculate the proportion of cases that were self-identifying gay men.

5.3.6 Phylogenetic analyses

I generated phylogenetic trees using the same procedures described in Chapter 4, section 4.4.5.3.

5.3.7 Ethics and information governance

Information governance advice and ethical approval for the analyses were sought from the PHE Research Support and Governance Office (RSGO). The RSGO approved the analyses as falling within public health surveillance and as such, no ethical approval was required. The

analyses were approved by the PHE Caldicott Panel in June 2017 (Appendix 4.2). Further details are provided in Chapter 4, section 4.4.6.

5.4 Results

5.4.1 Description of all WGS clusters

Between August 2015 and July 2017, there were 1,006 *S. flexneri* isolates with WGS data in England, of which 563 aggregated into 92 clusters defined at the 10-SNP threshold cut-off. Most clusters were classified as 'community' (n=36) or 'adult male' (n=36), followed by 'travel-associated' (n=14) and 'household' (n=6). However, when considered as a proportion of the overall number of cases across all clusters then the 'adult male' clusters accounted for most of the case burden (68.9%, n=388) followed by 'community' (22.4%, n=126), 'travel-associated' (6.4%, n=36) and 'household' (2.3%, n=13). The median cluster size was two cases (range: 2 to 240 cases) and the median cluster duration (i.e. the time between the first and last reported cases) was two months (range: 1 day to 24 months).

The median cluster size was three cases (range: 2 to 240 cases) for 'adult male' clusters, two (range: 2 to 13) for 'community', two (range: 2 to 8) for 'travel-associated' and two (range: 2 to 3) for 'household'. 'Adult male' clusters were generally larger; one-third (12/36) consisted of five or more cases compared to 14.3% (8/56) of other clusters (p=0.031). 19.4% (7/36) of 'community' clusters, 7.1% (1/14) of 'travel-associated' clusters and none (0/6) of the 'household' clusters consisted of five or more cases. The median cluster duration was five months (range: 1 day to 24 months) for 'adult male' clusters, two months (range: 1 day to 21 months) for 'community', 11 days (range: 1 day to 8 months) for 'travel-associated' and eight days (range: 1 day to 11 days) for 'household'. Cluster duration was longer for 'adult male' clusters compared to other clusters: Half (18/36) of 'adult male' clusters persisted for six months or longer compared to 25.0% (14/56) of other clusters (p=0.014) (36.1% [13/36] of 'community' clusters, 7.1% [1/14] of 'travel-associated' clusters and none [0/6] of the 'household' clusters persisted for six months or longer). There was one dominant 'adult male' cluster which consisted of 240 cases occurring over a 24-month period and 61.9% (240/388) of all 'adult male' cases were in this 10-SNP cluster.

5.4.2 Description of questionnaire data

The questionnaire data used in this chapter were described in detail in Chapter 4. Briefly, *S. flexneri* questionnaires were available for 190 cases, representing 37.8% (190/503) of all cases reported to GBRU from the HPTs participating in the pilot (42.4% in London, 28.3% outside London) and 21.9% (190/868) of all cases reported nationally during the pilot period (August 2015 to March 2017). 75.3% (n=143/190) of questionnaires were submitted by London HPTs. Self-reported sexual identity and recent sexual contact data were available for 88.9% (n=152/171) and 92.5% (n=123/133) of individuals with questionnaire data, respectively.

5.4.3 Clusters with linked WGS and questionnaire data

When using different SNP thresholds (0, 5, 10, 25), the size and distribution of clusters varied across the different cluster classifications (Table 5.1).

At the 10-SNP threshold (i.e. the standard for defining clusters at PHE), 34 clusters contained at least one case with a completed questionnaire, representing 37.0% (34/92) of all clusters that were detected nationally during the study period. Of these 34 clusters, 97.1% (33) belonged to clonal complex (CC) 245 while 2.9% (1) belonged to CC145 (see section 4.1.2.2 for a description of CCs). Most were ‘adult male’ (22/34), followed by ‘community’ (10/34) and ‘travel-associated’ (2/34) clusters. These 34 clusters contained a total of 401 cases, representing 71.2% (401/563) of all clustered cases reported nationally during the study period (August 2015 to July 2017). Clusters classified as ‘adult male’ accounted for most of the cases (86.8%, 348/401), followed by ‘community’ (10.7%, 43/401), and ‘travel-associated’ (2.5%, 10/401) clusters. 26.4% (106/401) of clustered cases at the 10-SNP threshold had linked questionnaire data.

Generally, as the SNP threshold increased (i.e. was relaxed) a higher number of cases clustered and clusters also became larger in size. At the 0-SNP threshold, most (62.5%, 15/24) clusters contained only two cases, whereas this proportion dropped to 37.0% (10/27) at the 25-SNP threshold.

Table 5.1: Size and distribution of clusters by SNP threshold cut-off and cluster classification

	25 SNP (27 clusters, 515 cases)			10 SNP (34 clusters, 401 cases)			5 SNP (39 clusters, 286 cases)				0 SNP (24 clusters, 69 cases)		
	Adult male	**Community**	**Travel**	**Adult male**	**Community**	**Travel**	**Adult male**	**Community**	**Travel**	**Household**	**Adult male**	**Community**	**Travel**
Number of clusters	10	14	3	22	10	2	28	8	2	1	20	3	1
Number of cases	424	76	15	348	43	10	246	28	10	2	51	11	7
Cluster size													
Median (range)	4 (2-313)	3 (2-22)	5 (2-8)	4 (2-240)	3 (2-8)	5 (2-8)	3 (2-144)	3 (2-8)	5 (2-8)	-	2 (2-5)	2 (2-7)	-
2	4	5	1	9	5	1	11	4	1	1	13	2	0
3-4	2	4	0	3	1	0	9	2	0	0	6	0	0
5-9	0	3	2	6	4	1	5	2	1	0	1	1	1
10-19	2	1	0	3	0	0	2	0	0	0	0	0	0
20+	2	1	0	1	0	0	1	0	0	0	0	0	0
Number (%) of cases with a questionnaire	106 (25.0)	18 (23.7)	3 (20.0)	89 (25.6)	15 (34.9)	2 (20.0)	74 (30.1)	11 (39.3)	2 (20.0)	2 (100.0)	25 (49.0)	3 (27.3)	1 (14.3)

5.4.4 SNP clustering validation

Overall sensitivity and specificity of the cluster classifications at a range of SNP thresholds are presented in Figure 5.2. Sensitivity (i.e. the proportion of MSM that belonged to an 'adult male' cluster) was high (≥94%) at all SNP thresholds whereas specificity (i.e. the proportion of non-MSM that belonged to a non-sexual transmission cluster) ranged from 68.8% (95% CI: 41.7% to 85.8%) at the 5-SNP threshold to 100% at the 0-SNP threshold but the confidence intervals at all thresholds overlapped and the number of cases included at the 0-SNP threshold was very small (n=3).

At the 10-SNP threshold, sensitivity was 95.1% (77/81 [95% CI: 88.0 to 98.1%]). At this threshold, some MSM cases belonged to a 'community' cluster (4.9% [4/81]) and all reported recent same-sex sexual contact (Table 5.2). Specificity was 72.2% (13/18, [95% CI: 49.1 to 87.5%]). Among the non-MSM cases, 27.8% (5/18; all were heterosexual men) belonged to an 'adult male' cluster, 61.1% (11/18) to a 'community' cluster and 11.1% (2/18) to a 'travel-associated' cluster. At the 25-SNP threshold, sensitivity was 100% (95% CI: 96.0% to 100%) and specificity was 77.8% (95% CI: 59.2 to 89.4%). Among the non-MSM cases, 22.2% (6/27, all heterosexual men) belonged to an 'adult male' cluster, 66.7% (18/27) to a 'community' cluster and 11.1% (3/27) to a 'travel-associated' cluster. For clusters at all SNP thresholds, there were adult men who did not provide information on sexual identity and/or recent sexual contact that clustered with men identifying as gay (Table 5.2).

For some clusters, the classification changed when applying different SNP thresholds (Table 5.3). Three 'community' clusters (two 10-SNP, one 5-SNP) were part of a larger 25-SNP 'adult male' cluster. These 'community' clusters did not meet the 'adult male' cluster criteria: for clusters comprising between two and five cases at least one case was female (e.g. Cluster ID 1 - M:F 1:1, 50% men aged 16 years or older) and for clusters comprising more than five cases, less than 90% were men aged 16 years or older (e.g. Cluster ID 2 - M:F 6:1, 85.7% men aged 16 years or older). All isolates in the 25-SNP 'adult male' cluster (highlighted in grey, Figure 5.3) belonged to the same phylogenetic clade for which 10-SNP clustering identified multiple discrete clusters (7 'adult male'; 2 'community') (Figure 5.4). The median pairwise SNP distance between any two isolates in the 25-SNP 'adult male' cluster was 21

(range 0 to 47). Overall, 74.2% (69/93) of isolates (clustered and non-clustered) from gay-identifying men belonged to this clade (Figure 5.3).

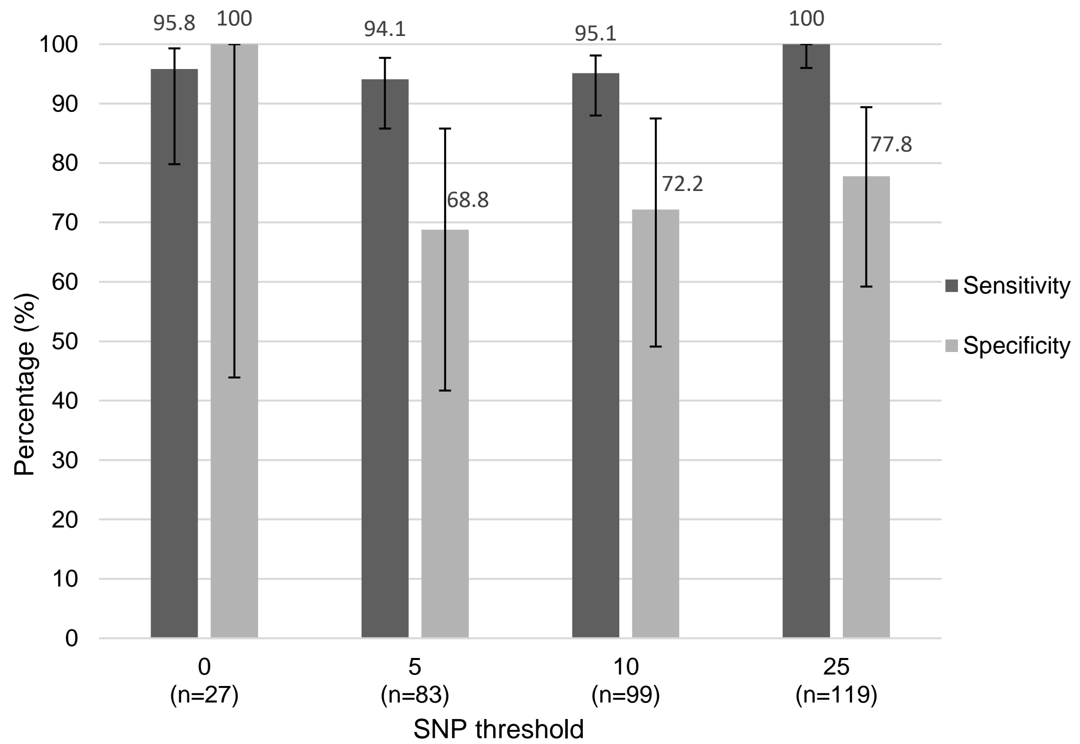


Figure 5.2: Sensitivity and specificity of the cluster classification tool at different SNP thresholds using the questionnaire data as the gold standard

Sensitivity represents the proportion of MSM that belonged to an 'adult male' cluster and specificity represents the proportion of non-MSM (heterosexual men, women and children under the age of 18 years) that belonged to a 'community', 'travel-associated' or 'household' cluster. Error bars represent 95% Confidence Intervals (CIs).

Table 5.2: Sexual identity, recent sexual contact and foreign travel for clustered cases with a completed questionnaire

	25 SNP (N=127)			10 SNP (N=106)			5 SNP (N=89)				0 SNP (N=29)		
	Adult male n=106	Community n=18	Travel N=3	Adult male N=89	Community N=15	Travel N=2	Adult male N=74	Community N=11	Travel N=2	Household N=2	Adult male N=25	Community N=3	Travel N=1
Sexual identity													
Gay men	90 (84.9)	0	0	75 (84.3)	4 (26.7)	0	63 (85.1)	4 (36.4)	0	0	23 (92.0)	1 (33.3)	0
Heterosexual men	6 (5.7)	3 (16.7)	2 (66.6)	5 (5.6)	1 (6.7)	1 (50.0)	5 (6.76)	0	1 (50.0)	1 (50.0)	0	0	1 (100.0)
Adult men [a]	10 (9.4)	0	0	9 (10.1)	0	0	6 (8.1)	1 (9.1)	0	0	2 (8.0)	0	0
Lesbian	0	0	0	0	0	0	0	0	0	0	0	0	0
Heterosexual women	0	11 (61.1)	0	0	9 (60.0)	1 (50.0)	0	6 (54.5)	0	1 (50.0)	0	2 (66.7)	0
Adult women [a]	0	2 (11.1)	1 (33.3)	0	0	0	0	0	1 (50.0)	0	0	0	0
Other (<18 years old)	0	2 (11.1)	0	0	1 (6.7)	0	0	0	0	0	0	0	0
Recent sexual contact [b]													
With man	65 (61.3)ǂ	0	0	56 (62.9)ǂ	4 (80.0)	0	45 (60.8)§	4 (80.0)	0	0	21 (84.0)	1 (100.0)	0
With woman	3 (2.8)	2 (66.7)	0	2 (2.2)	1 (20.0)	0	2 (2.7)	0	0	1 (100.0)	0	0	0
Gender not disclosed	2 (1.9)	0	0	2 (2.2)	0	0	1 (1.4)	1 (20.0)	0	0	0	0	0
No sexual contact	30 (28.3)	1 (33.3)	2 (100.0)	24 (27.0)	0	1 (100.0)	21 (28.4)	0	1 (100.0)	0	3 (12.0)	0	1 (100.0)
Not known	6 (5.7)	0	0	5 (5.6)	0	0	5 (6.8)	0	0	0	1 (4.0)	0	0
Foreign travel [c]													
Europe	11 (10.4)	0	0	9 (10.1)	1 (6.7)	0	7 (9.5)	1 (9.1)	0	0	2 (8.0)	1 (33.3)	0
Caribbean	1 (0.9)	0	0	1 (1.1)	0	0	1 (1.4)	0	0	0	0	0	0
Middle East	2 (1.9)	1 (5.6)	1 (33.3)	2 (2.2)	0	1 (50.0)	1 (1.4)	0	1 (50.0)	0	0	0	1 (100.0)
North Africa	0	2 (11.1)	0	0	1 (6.7)	0	0	1 (9.1)	0	0	0	0	0
South Asia	0	4 (22.2)	1 (33.3)	0	1 (6.7)	1 (50.0)	0	0	1 (50.0)	0	0	0	0
Sub-Saharan Africa	1 (0.9)	4 (22.2)	0	1 (1.1)	3 (20.0)	0	1 (1.4)	3 (27.3)	0	0	0	1 (33.3)	0
No or unknown recent travel	91 (85.9)	7 (38.9)	1 (33.3)	76 (85.4)	9 (60.0)	0	64 (86.5)	6 (54.5)	0	2 (100.0)	23 (92.0)	1 (33.3)	0

a. Sexual identity not reported. b. Denominator includes adult men (≥18 years) only. c. Foreign travel as recorded on questionnaire; data missing for 1 case. ǂIncludes 2 men (§1 man) who did not disclose their sexual identity (i.e. their sexual identity is reported here as "adult men") but who reported recent same-sex sexual contact.

Table 5.3: Impact of SNP threshold cut-off on the size, distribution and classification of the cluster

Cluster ID	SNP threshold	SNP address	Classification	Total cases	M:F ratio	No. (%) adult men	No. (%) gay-identifying men±	No. (%) Foreign travel	Min SNP	Max SNP	Median SNP
1	25	34.42.42.42.%	Adult male	313	301:8	301 (97.4)	69/81 (85.2)	20 (6.4)	0	47	21
	10, 5, 0	34.42.42.42.537.634.796	Community	2	1:1	1 (50.0)	1/1 (100)	0	0	0	0
2	25	34.42.42.42.%	Adult male	313	301:8	301 (97.4)	69/81 (85.2)	20 (6.4)	0	47	21
	10	34.42.42.42.344.%	Community	7	6:1	6 (85.7)	3/3 (100)	0	1	11	5
	5	34.42.42.42.344.396.%	Community	5	4:1	4 (80.0)	3/3 (100)	0	1	7	4
3	25	34.42.42.42.%	Adult male	313	301:8	301 (97.4)	69/81 (85.2)	20 (6.4)	0	47	21
	10	34.42.42.42.42.%	Adult male	240	232:5	232 (97.9)	48/56 (85.7)	15 (6.3)	0	38	17
	5	34.42.42.42.42.526.%	Community	4	3:1	3 (75.0)	0/1*	0	2	7	4
4	25	4.49.49.129.%	Community	10	4:5	4 (44.4)	0/2	3 (30.0)	3	20	13
	10	4.49.49.129.325.%	Community	8	2:5	2 (28.6)	0/2	2 (25.0)	3	17	11
	5	4.49.49.129.324.663.%	Household	2	1:1	1 (50.0)	0/2	0	3	3	3

*Adult man - did not report sexual identity and preferred not to disclose gender of partner. M:F male to female sex ratio. ± Denominator includes individuals with a questionnaire only. Foreign travel as reported through questionnaires or laboratory records (for those without a questionnaire). Min (minimum), Max (maximum) and Median SNP represent the pairwise SNP distances between any two isolates in a cluster, corrected for recombination, to demonstrate the level of genetic variation

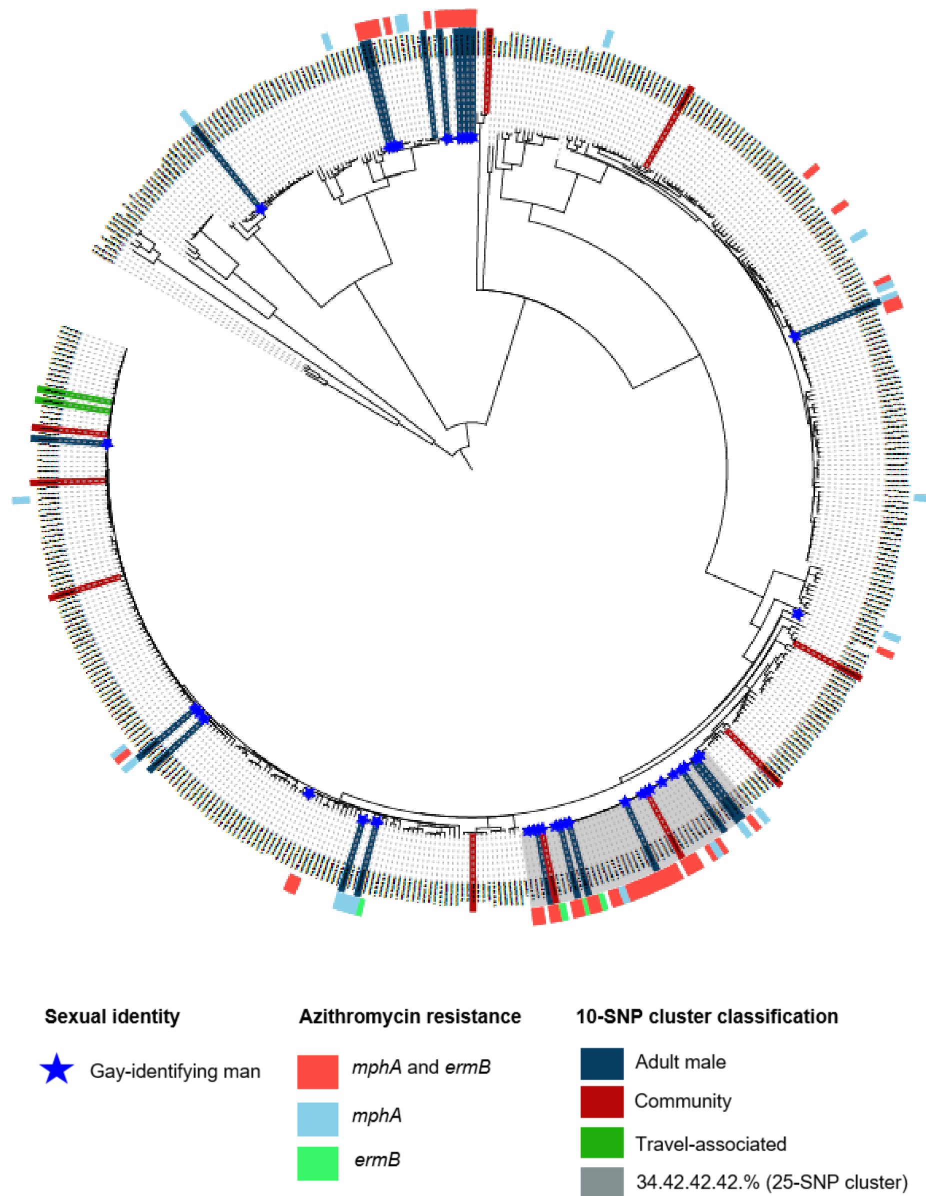


Figure 5.3: Phylogeny of *S. flexneri* CC245 showing the distribution of clusters, genotypic markers of azithromycin resistance and gay-identifying men

Mid-point rooted maximum likelihood phylogenetic tree showing a single representative isolate from each 10-SNP cluster (N=474) for CC245 during the study period (N=926) and seven reference strains for each phylogenetic group.[297] The number of isolates represented by each branch tip ranges from one to 240. Clusters (i.e. two or more cases) containing at least one case with a questionnaire are coloured at the tips according to cluster classification (33 clusters representing 394 cases). Genotypic markers of azithromycin resistance are shown as a coloured track on the outside of the tree. Isolates from gay-identifying men are shown as blue stars. An 'adult male' 25-SNP cluster is highlighted in grey.

10-SNP clusters

- 34.42.42.42.285.% (adult male)
- 34.42.42.42.291.% (adult male)
- 34.42.42.42.344.% (community)
- 34.42.42.42.356.% (adult male)
- 34.42.42.42.371.% (adult male)
- 34.42.42.42.378.% (adult male)
- 34.42.42.42.42.% (adult male)
- 34.42.42.42.440.% (adult male)
- 34.42.42.42.537.% (community)

Age group

- <18
- 18-24
- 25-34
- 35-44
- 45+

Gender

- Male
- Female

Sexual risk

- Gay man
- Reported sex with a man
- Heterosexual man
- Not known

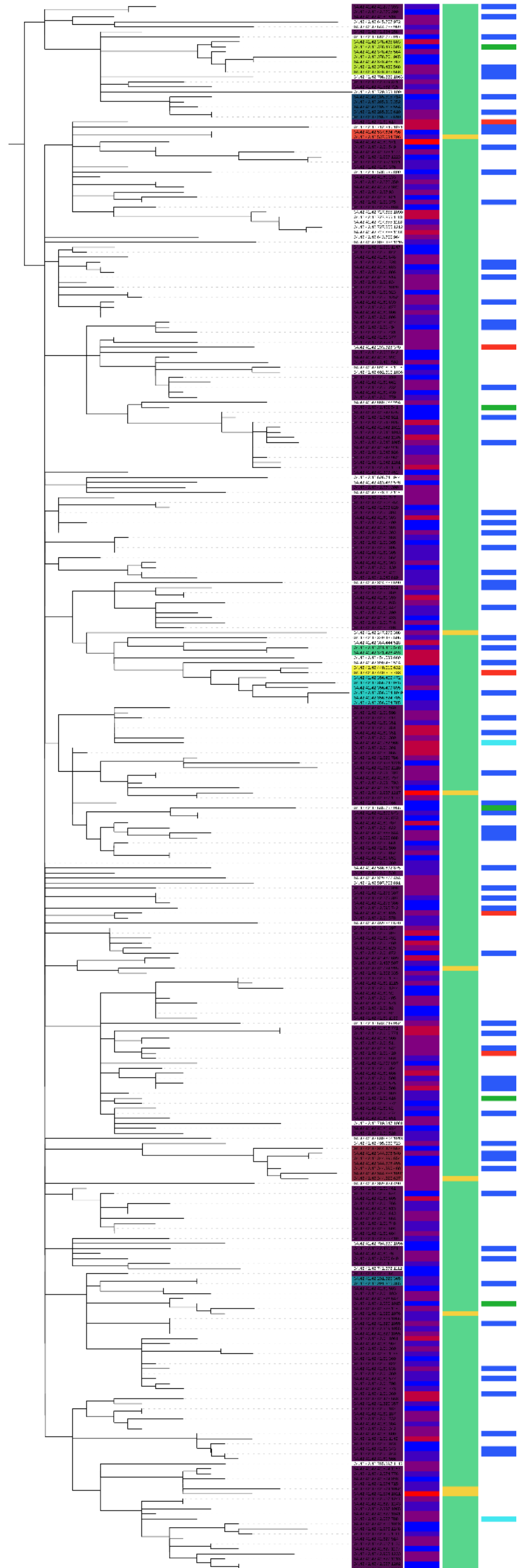


Figure 5.4: Phylogeny of 25-SNP ‘adult male’ cluster 34.42.42.42.%

Mid-point rooted maximum likelihood phylogenetic tree containing all isolates belonging to an ‘adult male’ 25-SNP cluster (34.42.42.42.%, N=313). Isolates are labelled by SNP address. 10-SNP single linkage clusters containing at least one case with a questionnaire are coloured at the tips by cluster (9 clusters, 273 cases in total). Branch tips that are not coloured represent cases that did not cluster with another case at the 10-SNP threshold (n=33 cases), or 10-SNP single linkage clusters that did not contain a case with a questionnaire (two ‘adult male’ clusters, seven cases in total). Selected demographic data and sexual risk are represented as coloured strips. Sexual risk data are presented for cases with a questionnaire.

5.4.5 Genotypic markers of azithromycin resistance

Genotypic markers of azithromycin resistance were detected in the genomes linked to 84.5% (294/348; 29 *mphA*, 11 *ermB* and 254 *mphA* and *ermB*) of cases belonging to 'adult male' clusters and 20.9% (9/43; all *mphA* and *ermB*) of cases belonging to 'community' clusters at the 10-SNP threshold (Figure 5.3). Of cases that had genotypic markers of azithromycin resistance and where a questionnaire was completed, 83.3% (65/78) and 100% (4/4) of isolates within 'adult male' and 'community' clusters respectively were gay-identifying men. The remaining 16.7% (n=13) of cases within 'adult male' clusters comprised five heterosexual-identifying men, two of whom reported recent sexual contact with a woman, and eight adult men who did not provide information on sexual identity, although two reported recent sexual contact with a man. Four gay-identifying men belonged to two 'community' clusters that were part a larger 'adult male' cluster at the 25-SNP threshold. These clusters were described earlier in section 5.4.4 and in Table 5.3.

5.5 Discussion

5.5.1 Key findings

In this chapter, I provide robust evidence that SNP clustering using WGS data linked to basic demographic data available on laboratory report forms is an effective means for distinguishing clusters likely associated with sexual transmission in MSM from other, non-sexual transmission clusters in near real-time. This approach therefore seems suitable for use by PHE to inform the delivery of rapid, targeted and appropriate public-health responses that aim to prevent onward transmission from infected cases and control the spread of *S. flexneri.*

5.5.2 Strengths and limitations

This is the first time that robust data on sexual identity and behaviour have been available for a large sub-set of cases in England, and I have used these data as the gold standard to validate a real-time SNP clustering algorithm for classifying transmission clusters. My cluster analysis used a large, representative sample of sequenced isolates that were submitted to the national reference laboratory over a two-year period, improving the generalisability of my

findings to other settings and countries. I have demonstrated that the SNP typing tool provides a fast and accurate method of classifying transmission clusters in the absence of questionnaire data, which is particularly valuable in regions where the collection of exposure information is challenging. Furthermore, the rapid synthesis and analysis of WGS data might facilitate rapid and appropriate public health action.

The main limitation of this study was that behavioural data were only available for HPTs that were selected to participate in a national pilot study. These pilot HPTs consisted of geographical regions that have larger MSM populations, including London, Manchester and Brighton and Hove.[298] As a result, my validation analysis comprised of a higher proportion of 'adult male' clusters when compared to the national dataset. Furthermore, a higher proportion of all clustered cases included in my analysis belonged to the 'adult male' clusters. The overrepresentation of 'adult male' clusters and cases in my analysis could have biased the cluster validation analysis if the HPTs were more likely to follow-up MSM cases, or if the HPTs not included in the pilot were more likely to see 'adult male' clusters which were linked through non-sexual transmission. An additional limitation was that sexual identity and recent sexual behaviour were only collected for people aged 18 years or older, whereas the 'adult male' cluster classification used a cut-off of 16 years or older, however, the impact on the validation analysis is negligible due to the small number of individuals affected.

As discussed in Chapter 4, the study was restricted to cases who presented to healthcare, with a stool sample collected for investigation by local hospital laboratories and referral to the GBRU for molecular typing; approximately two thirds of *Shigella* spp. isolated at local hospital laboratories are referred to the GBRU.[314] It is important to recognise that the clusters were classified based on the characteristics of those who had their isolates referred to the GBRU, and it is not clear how representative these are of the wider transmission cluster. A range of factors are likely to have influenced whether an isolate was included, such as patterns of health-seeking behaviour, testing practices and gender disparities in travel, food consumption or childcare.[201] Furthermore, an analysis of *Shigella* spp. surveillance data from the USA found that the odds of severe shigellosis were higher for men compared to women among adults aged 18-49 years old. However, the role of sexual behaviour was not assessed.[303] Collectively,

such factors are more likely to lead to an overestimation in the proportion of cases transmitted through sexual contact.

5.5.3 Interpretation of clustering data

In this study, I found that most cases belonged to 'adult male' clusters and that these clusters consisted largely of MSM (where questionnaire data were available) with no or unknown recent travel history. One large 'adult male' 25-SNP cluster persisted for 24 months, consistent with ongoing sexual exposure and potential for reinfection within large and dense sexual networks.[4,112,117,315]

Using a range of SNP-thresholds, there were some clusters that consisted of both MSM and non-MSM (heterosexual men, women and children), although the number of non-MSM cases was small providing evidence that transmission beyond sexual networks is probably limited. There was one 0-SNP cluster consisting of a gay-identifying man and an adult woman. Although these two sequences were genetically indistinguishable, care is needed in the interpretation because the WGS data do not provide evidence of direct transmission or indicate the direction of transmission, and cannot exclude any intervening cases.

Of concern, and in contrast to non-sexual transmission clusters, genotypic markers of azithromycin resistance were detected in most cases belonging to 'adult male' clusters. High-level azithromycin resistance in MSM transmission networks has been shown to play a role in driving the spread of *Shigella* spp.,[24] and this raises concerns about future treatment options for shigellosis and the wider global problem of AMR.

5.5.4 Public health implications

GBRU typing data are used for routine surveillance and are currently assessed on a weekly basis for active clusters at the 10-SNP threshold. The decision to investigate a cluster depends on several factors including cluster classification, duration, size and geographic spread. Cluster-specific interventions are not usually implemented for 'adult male' clusters that are geographically and temporally dispersed. Instead, public health messages targeted towards MSM and clinicians nationally aim to raise awareness of shigellosis and offer advice on what

men can do to lower their risk e.g. washing hands and genitals before and after sex, using condoms for risk practices and avoiding the use of shared sex toys.[26] Despite this, transmission continues.

My findings can inform best practice for *Shigella* spp. cluster investigations in England, and will be of interest to any country developing and implementing WGS for surveillance and/or outbreak detection and response. However, this methodology is dependent on building a comprehensive library of WGS data against which to classify new cases. My analysis showed that real-time WGS and analysis of *S. flexneri* isolates at the 10-SNP threshold can be used to detect likely clusters of sexual transmission in MSM with a high level of accuracy to inform the public health response (sensitivity: 95.1% [95% CI: 88.0 to 98.1%], specificity: 72.2% [95% CI 49.1 to 87.5%]). In practice, however, restricting the SNP threshold to 10-SNPs may lead to investigation of clusters that appear to be part of different sexual networks, or result in MSM not receiving health advice about preventing the sexual transmission of infection. While the 10-SNP threshold provides a good default for cluster detection, a flexible approach is important to fully understand the underlying transmission network and likely transmission route, particularly in clusters of MSM comprising fewer than 10 cases where at least one is female. By changing the SNP threshold and exploring the phylogenetic context of isolates, I found small 'community' clusters consisting of MSM and adult women, that were part of a larger 25-SNP 'adult male' cluster. Persistent transmission networks sustained through sexual contact over longer periods of time will naturally exhibit greater genetic diversity. As a result, a more relaxed threshold (such as 25-SNPs) should be considered when investigating these clusters.

The SNP clustering tool I validated in this study provides a potential mechanism to provide appropriate and targeted health care advice to cases, not only to MSM who are unaware of the sexual transmission of *Shigella* spp., but to men who may not identify as gay or bisexual, or reveal same-sex sexual contact, but are likely to be part of an MSM network, and therefore at higher risk of acquiring *Shigella* spp., as well as other STIs and HIV infection. Early identification of sexual transmission could be used to ensure that those affected are referred promptly to SHCs for further STI testing, investigations and interventions. Using WGS data to inform patient management raises important ethical issues with regards to confidentiality and

privacy of information, as well as the potential for deductive disclosure of sexual partners and networks that could lead to discrimination.

The utility of SNP clustering for improving case ascertainment during outbreak investigations and revealing linked cases not identified prior to the implementation of WGS has been described previously.[284,288,316] My analysis demonstrates that the SNP clustering tool can be used to detect and classify different transmission clusters effectively in the absence of questionnaire data. Nonetheless, the use of a questionnaire to collect detailed information on demographics and potential exposures could help to unravel the characteristics of the cluster, improve confidence in understanding about how transmission might be occurring and strengthen the public health response. For example, epidemiological data may identify people that share a common source such as attendance at a specific sex-on-premises venue and these clusters may be amenable to more specific and targeted public health action including improvements in hygiene and the provision of condoms. The addition of epidemiological data to SNP clustering data may also reveal novel links between MSM and other populations in the community, particularly where MSM are within a known risk group (e.g. food handlers or healthcare workers). Transmission between MSM and non-MSM population groups, although currently limited to small numbers, should not be ignored. Sustained transmission in sexual networks could seed community-acquired outbreaks, which is a concern given the high rates of antimicrobial resistance in the MSM population, and so these clusters should be prioritised for public health action.

5.5.5 Chapter summary

In this chapter I aimed to validate a real-time SNP clustering algorithm for distinguishing clusters of *S. flexneri* in sexual networks of MSM from non-sexual transmission clusters to inform the public health response. In summary:

- Using questionnaire data as the gold standard, I have shown that 'adult male' clusters represent likely sexual transmission in MSM. This shows that real-time SNP typing in the absence of questionnaire data can be used to detect and distinguish clusters representing likely sexual transmission in MSM from other, non-sexual transmission

clusters and might enable rapid delivery of a more appropriate and targeted public health response.

- Investigating clusters using a standard threshold of 10-SNPs is a reasonable default approach. However, flexibility is required in terms of genetic relatedness (i.e. SNP threshold) to provide a deeper understanding of the genetic diversity within a transmission network and the potential route of infection.
- A high proportion of isolates belonging to 'adult male' clusters harboured genotypic markers of azithromycin resistance, indicating that azithromycin resistance in clusters of *S. flexneri* is a marker for sexual transmission in MSM.

Chapter 6: Discussion

In this chapter, I summarise my PhD research and discuss the key findings within the context of the wider clinical and public health implications. I also consider future avenues of investigation that could address unanswered questions arising from my research.

6.1 Summary of PhD research

The aim of this PhD research was to investigate and describe the clinical and epidemiological characteristics, risk factors and burden of infection associated with bacterial enteric pathogens (BEPs) in MSM that could inform the development, targeting and delivery of more appropriate and effective interventions.

My objectives were to:

1. Provide robust and representative estimates of prevalence to assess the burden of BEPs in MSM
2. Investigate the potential for subclinical infection or asymptomatic carriage in sustaining transmission of BEPs in MSM
3. Describe the clinical, behavioural and epidemiological risk factors of BEPs in MSM
4. Explore the association between BEPs in MSM and genotypic markers of antimicrobial resistance, and how this relates to a previous bacterial STI diagnosis
5. Improve identification and characterisation of sexual versus other types of transmission of BEPs

Following the literature review in Chapter 2 to provide context for my PhD research, Chapter 3 describes a cross-sectional study of 2116 MSM attending a large SHC in central London to investigate the prevalence of BEPs and the associated risk factors. I found that 1 in 10 MSM had a BEP detected, and detection was associated with a suite of higher-risk sexual behaviours. Among MSM who had a BEP detected, the presence of a genotypic marker for azithromycin resistance was associated with a previous STI diagnosis. In Chapter 4, I used molecular epidemiology to identify and characterise transmission of *Shigella flexneri* in sexual networks of MSM and explored how this relates to non-sexual transmission within the

community. I described two domestically circulating MSM clades of *S. flexneri* (serotypes 2a and 3a) which were associated with genotypic markers of azithromycin resistance. Over one third of cases linked to these clades were from people living with HIV, indicating the overlap between these epidemics. There were strikingly different epidemiological and molecular characteristics between MSM and non-MSM cases, with evidence of sustained transmission through sex between men and limited transmission between these groups. In Chapter 5, I validated a real-time SNP clustering algorithm for identifying and distinguishing clusters of *S. flexneri* cases representing sexual transmission in MSM from clusters of cases representing non-sexual transmission which might enable the deployment of rapid and appropriate public health responses to control the spread of *Shigella* spp.

6.2 Clinical and public health implications

6.2.1 Spread and persistence of BEPs in MSM

As discussed in Chapter 2, the spread and persistence of sexually transmissible pathogens in a given population are determined by three main components: the rate of partner change, the duration of infectiousness and the transmission probability.[217]

In Chapter 3, I found that BEP detection was associated with a suite of higher-risk sexual behaviours, including that MSM who reported a higher number of new sexual partners in the past three months were more likely to have a BEP detected than those with fewer partners. This observation provides good circumstantial evidence of sexual transmission and indicates that, as with other STIs, the rate of partner change is an important component in the spread of BEPs in MSM. At a population level, the rate of partner change influences the number of new infections that an infected person can generate and thereby the growth of the epidemic.[219] However, the chances of an infection spreading rapidly and persisting within a population are also dependent on patterns of sexual mixing and the structure of the sexual network.[218-220] STIs spread more rapidly in networks where concurrent (as opposed to serial monogamous) partnerships predominate. In addition, the chances of an infection persisting are greatest when those who report the highest risk sexual behaviours (including higher numbers of new sexual partners) preferentially choose partners with similar behaviours. The combination of high rates

of partner change, concurrency, and assortative mixing can create tightly connected sexual networks where the STI is highly concentrated within specific subgroups of the population while the prevalence in the general population is low.[218,219] In this study, I found that high risk sexual behaviours were correlated with each other (Chapter 3), suggesting that BEPs are likely concentrated within specific sub-groups of the MSM population that are highly sexually active. My analysis of the transmission of *S. flexneri* among sexual networks of MSM (Chapter 4) also supports this idea, as it shows that transmission is primarily associated with the spread of two dominant clades, of which over one third (37.4%) of isolates were from people (mainly MSM) living with diagnosed HIV. Furthermore, 76.7% (240/313) of cases within the largest MSM clade fell within the same 10-SNP cluster that persisted over a two-year period (Chapter 5). The low level of genetic diversity and prolonged persistence of this cluster demonstrates the large number of cases that were linked by recent transmission events within a dense sexual network.

The duration of infectiousness is influenced by the biology of the infecting pathogen and its interaction with the host immune system and the likelihood of developing symptoms and seeking treatment.[218,317] For BEPs, most infections typically lead to clinical symptoms after a short incubation period and people are most likely to be infectious when they are symptomatic, since gastrointestinal symptoms like diarrhoea are reflective of a higher pathogen load.[33,317] Treatment is one way to shorten the infectious period and thereby control infection at a population level (although an important exception are *Salmonella* spp., where antimicrobial treatment may prolong faecal shedding).[63,318] After symptoms have resolved, an individual is infectious for as long as the organism is excreted in stools, which can be up to several weeks for BEPs without treatment.[319] For instance, *Shigella* spp. can be excreted for four weeks after symptoms resolve, although longer term excretion for many months has been reported,[320,321] and *Campylobacter* spp. can be excreted for two to three weeks, and possibly up to two months, post infection.[48] Asymptomatic carriage or subclinical infection may provide a reservoir to enable the pathogen to persist in the population because those affected may not change their sexual behaviour whilst harbouring the pathogen for long periods.[33] In Chapter 3, I found that nearly one in 10 MSM attending the UK's largest SHC had a BEP detected and most of these men at the time of testing had no symptoms of gastroenteritis, which is

consistent with the theory that asymptomatic carriage might be playing an important role in sustaining transmission in this population. As discussed in Chapter 3 (see section 3.5.2), because I used an opt-out study design with routine data collection, I may have underestimated the proportion of MSM who had gastrointestinal symptoms. In addition, the use of routine data meant that I was unable to explore the type and duration of symptoms in further depth, which could have helped with interpretation.

Quantifying the probability of transmission between an infected and susceptible person is challenging, particularly for STIs, and is influenced by both biological and behavioural factors. The infectious dose of the pathogen is also likely to be important.[317] Some BEPs can cause infection with as little as 10 organisms (e.g. *Shigella* spp.). Combined with stability outside of the body on fomites, a low infectious dose can mean that the probability of transmission is high. On the other hand, *Salmonella* spp. have a higher infectious dose and so people need to ingest larger volumes of faecal matter for infection to occur, which might explain why the pathogen was not detected among MSM in my study (Chapter 3). The type of sexual practice will also influence transmission risk, for example, direct oral-anal contact (i.e. rimming) is likely to result in exposure to a larger number of organisms, compared to other sexual activities, but likely occurs alongside other sexual risk behaviours such as having a higher number of new sexual partners, recreational drug use and group sex; I was unable to collect specific information about these sexual practices due to the opt-out study design. Furthermore, population immunity may build as an infection spreads, which could decrease the overall pool of susceptible people in a given population.[219,317] This means that infected people are less likely to meet susceptible people, even when there is high-risk sexual behaviour in a population.[317] The rate of transmission may start to rise again with a supply of newly susceptible people into a sexual network at risk of exposure or through waning levels of immunity in the existing exposed population.[154,219,317] For some BEPs, such as *Campylobacter* spp.[322] and *Shigella* spp.,[154] there is evidence of protective immunity that can prevent clinical disease following exposure to homologous strains. However, immunity may not protect against subsequent colonisation.[322] Moreover, antibody levels likely decline over time.[155]

Enteric pathogens are typically considered short duration, high transmission probability pathogens. To thrive in a given population, they likely require a high rate of contact within a dense population structure characterised by high rates of partner change and other risk behaviours such as rimming, fisting and group sex, likely facilitated by chemsex. A similar network structure is necessary for the persistence of bacterial STIs like gonorrhoea, which is usually short-lived (particularly when infections are symptomatic and treated) with a high transmission probability.[219] In Chapter 3, BEP detection was associated with a concurrent and previous bacterial STI diagnosis, suggesting that bacterial STIs and enteric pathogens occur in the same sexual networks.

The patterns of sexual behaviour and risk that influence the spread of sexually transmissible pathogens can change over time.[317] For instance, the amount of time that a person spends as a member of a high-risk sexual network can have a bearing on the number of new infections and overall prevalence.[323] Pines *et al.* (2014) assigned sexual behaviour risk scores to HIV-negative MSM participating in a cohort study of HIV in the United States and found variations in the pattern of behavioural risk over time.[324] Thus, higher-risk sexual networks are highly dynamic, where people enter as they start engaging in higher-risk sexual behaviours and leave as they change their sexual behaviour.

To summarise, the spread and persistence of BEPs in MSM is complex and is influenced by the underlying prevalence of infection, levels of pre-existing immunity and patterns of sexual risk behaviour, including high rates of partner change, within highly dynamic sexual networks. The high probability of transmission per sex act and the duration of infectiousness of BEPs also determines their ability to spread, and asymptomatic carriage could provide a reservoir that enables the infection to persist.

6.2.2 Prevention and control of enteric pathogens in MSM

6.2.2.1 Population awareness

One way to control the transmission of BEPs in MSM is to raise awareness of the potential for sexual transmission of BEPs and the symptoms they can cause, as well as ways to prevent infection. My analysis of *S. flexneri* cases found that only 36% of MSM had previously heard

of shigellosis, indicating that health education remains a major challenge (Chapter 4). Similar findings have been reported in other studies. A qualitative study of MSM in four cities across England in 2015 exploring perceptions and attitudes towards a range of STIs found that most MSM had never heard of shigellosis and in a ranking exercise, it was considered of little concern when compared against other STIs and HIV.[325] Furthermore, a survey of MSM attending three SHCs in London in 2016 found that only 29% of MSM had heard of shigellosis.[158] More recently, an anonymous online survey of 3646 MSM recruited through dating apps in 2017 found that 26.6% had heard of shigellosis, but actual knowledge about how the pathogen is spread and related symptoms was lower (16.5%).[258]

To date, targeted public health campaigns have attempted to raise awareness of shigellosis and hepatitis A among MSM, and the preventative measures that can be taken to avoid infection, by displaying health information in SHCs and on social networking sites. These campaigns have attempted to promote hygiene measures before and after sex, the use of latex barriers for activities such as rimming or fisting, refrainment from sharing sex toys and abstinence from sexual contact for seven days after symptoms have resolved. My research shows that a range of BEPs are circulating among sexually active MSM (Chapter 3) and efforts should therefore focus on developing proactive messaging about the broader risk of enteric pathogens spread through sexual contact. There are challenges associated with this given the potential for asymptomatic carriage, which would need to be incorporated into any health messages (see section 6.2.2.5).

6.2.2.2 Targeted interventions

In Chapter 3, I found some sub-groups of MSM were more likely to have a BEP detected and these risk factors might be used to target interventions, including health education and population awareness, to people at risk of BEPs. Appropriate targeting and outreach to people that are at greatest risk of acquiring and transmitting infections is a fundamental concept that underpins traditional STI control programmes. Interrupting transmission in high-risk groups may be sufficient to reduce transmission in the wider population. SHCs are a suitable environment in which to identify those who may be at greatest risk as my study shows that these pathogens are circulating among MSM attending a London sexual health clinic. In

addition, semi-structured interviews conducted with 34 MSM diagnosed with *S. flexneri* serotype 3a during a UK outbreak found that nearly all (94%) had attended a SHC or HIV service, and 69% had attended in the three months prior to their shigellosis diagnosis for routine testing and/or HIV care.[4]

MSM living with HIV and HIV-negative MSM attending SHCs for HIV-PrEP were more likely to have a BEP detected compared to HIV-negative MSM not taking HIV-PrEP (Chapter 3), and these MSM could be asked about gastrointestinal symptoms and given specific health education messages about enteric pathogens, the behaviours that are likely to facilitate transmission and preventative measures to avoid infection. The association between BEP detection and sexual behaviour was less clear in the sub-group of MSM living with HIV (Chapter 3). Nonetheless, over one third (37.4%) of *S. flexneri* isolates that belonged to the domestically circulating MSM clades were from people living with HIV, highlighting the need to raise awareness about the risk of enteric pathogens in these men.

Recent evidence from an online survey of MSM suggested that engagement in behaviours that increase the likelihood of STI and BEP transmission varies by STI knowledge. Among HIV-negative/unknown status MSM, those with a good level of knowledge about STIs were more likely to report a higher number of sexual partners and recreational drug use prior to sex, but were less likely to report condomless anal sex, compared to those with a poor level of knowledge. Among MSM living with HIV, those with a good level of knowledge were more likely to report condomless anal sex than those with a poor level of knowledge. Furthermore, engagement in STI risk behaviours was higher generally among MSM living with HIV compared to HIV-negative/unknown status MSM even though they had a higher level of knowledge.[258] As such, controlling transmission ultimately requires a deeper understanding of the social and psychological motivations for engaging in higher-risk behaviours.

6.2.2.3 Clinical management

In Chapter 4, I found that most MSM (78.3%) infected with *S. flexneri* sought healthcare from their GP and/or a non-specialist hospital service such as A&E, with only one in five attending a SHC. Comprehensive management of clinical cases of enteric infections requires an awareness about the range of pathogens that could be responsible for gastrointestinal

symptoms and the potential route of transmission (i.e. sexual or non-sexual). This raises challenges for community healthcare settings such as general practice and A&E because clinicians may not routinely ask about sexual identity or sexual behaviours.[326] Furthermore, some patients may not wish to disclose their sexual history in non-specialist settings, either because of concerns about social stigma or because they do not consider it relevant.[326] Either way, not knowing the route of infection is likely to be sexual transmission could inhibit contact tracing and result in missed opportunities to prevent onward transmission and/or subsequent re-infection. Probable diagnoses of BEPs, particularly among adult men where there is no recent foreign travel history, should prompt clinicians to take a sexual history.[327] In late 2017, NHS England released a new standard recommending that all health services ask about sexual orientation. Although the standard provides a mechanism for healthcare providers to routinely ask this information, it is not mandatory.[328,329] Nonetheless, sexual identity (88.9%) and recent behaviour (92.5%) were well completed in my analysis of *S. flexneri* cases (Chapters 4 and 5), indicating that these questions are acceptable to patients. Improving awareness among healthcare providers in non-specialist settings about the possibility of sexual transmission of enteric pathogens is necessary. One way of doing this, in the first instance, could be to target GPs that have a special interest in sexual health and to conduct awareness campaigns in collaboration with the Royal College of Emergency Medicine, who represent A&E clinicians.

Among MSM attending a London SHC, BEP detection was associated with a concurrent or recent bacterial STI diagnosis (Chapter 3). Furthermore, I found that over one third (149/398) of *S. flexneri* isolates that fell phylogenetically within MSM clades were from people living with HIV. Given the overlap between enteric pathogens, STIs (such as gonorrhoea, syphilis and LGV) and HIV, the clinical management of enteric pathogens in MSM will likely require collaboration between different health services to ensure appropriate STI testing, clinical management and partner notification. In addition, sexual partners of those diagnosed with an enteric pathogen or STI should be notified and informed of their potential exposure and offered appropriate advice. Although gastrointestinal illness occurs frequently in people living with HIV, even among those who are taking treatment and have normal CD4 cell counts,[330] it is essential that microbiological examinations are performed to confirm whether symptoms are due to an

enteric pathogen. People who have advanced immunosuppression due to HIV infection or other co-morbidities are likely to be at greater risk of developing complications.[26]

My analysis of *S. flexneri* cases suggested that MSM experience more severe symptoms and clinical outcomes compared to other adults (i.e. heterosexual men and women) (Chapter 4). The mechanisms driving this are unclear and need to be explored in depth among a larger sample of cases who have detailed information on clinical symptoms, antimicrobial treatment and co-infections such as bacterial STIs. The use of a symptom severity score that incorporates duration and frequency for each specific symptom rather than a binary yes/no response could also be beneficial. Nonetheless, my findings emphasise the need to ensure MSM receive appropriate preventative advice and clinical care. Previous studies have reported that MSM are disproportionately affected by a wide range of poor health outcomes across mental, physical and sexual health.[331] A holistic approach that promotes the broader health and wellbeing of MSM, as well as HIV/STI and enteric infection prevention and management, is essential.[115]

6.2.2.4 Public health management

The public health management of enteric infections requires effective communication between multi-sectoral organisations. In England, a good example is provided by South East HPT, where the team has developed a collaborative relationship with their local SHCs to enable prompt identification of new enteric infections, and to ensure that public health follow-up is conducted as soon as possible after diagnosis, and preferably on the same day. People may be most receptive to advice when they are still symptomatic and most likely to understand why public health follow-up is necessary (M Courtney, personal communication, February 2019). Effective communication between HPTs and community healthcare providers is also likely to ensure that cases are referred for STI and HIV testing and appropriate clinical sexual health care. This points to a more holistic approach to the public health management of enteric pathogens which should be implemented nationally. The value of questionnaire completion as part of public health follow-up is discussed in section 6.2.4.

6.2.2.5 *Asymptomatic and sub-clinical infection*

Nearly one in 10 MSM attending a London SHC had a BEP and most of these MSM did not have any symptoms of gastroenteritis or diarrhoeal illness (Chapter 3). The presence of sub-clinical infection and asymptomatic carriage may prevent effective control of BEPs in MSM as infections remain unrecognised by cases such that infectious organisms may be passed onto others. Unlike traditional STIs where early diagnosis and treatment is clinically indicated for the individual, and at a population level acts to reduce the duration of infectiousness and hence the incidence of infection, treatment is not recommended for most cases of acute gastroenteritis, which is usually self-limiting. Furthermore, asymptomatic screening is not recommended, even in higher-risk populations, because the clinical implications of asymptomatic carriage and risk of onward transmission are unclear, and there is a risk of fostering AMR. Interrupting the spread of enteric pathogens in MSM therefore relies on behaviour change and contact tracing, however, high partner turnover and anonymous sexual partners represent significant challenges.[26] It is unlikely that asymptomatic screening would be recommended given the concerns highlighted but further understanding of the potential for onward transmission from asymptomatic carriers would be valuable to determine the impact on transmission at a population level.

6.2.3 Antimicrobial use and resistance in MSM

My study is the first to explore the prevalence of *mphA* in a large population of predominantly asymptomatic MSM, and how this relates to BEP detection and STI history (Chapter 3). I found that the prevalence of *mphA*, a genotypic marker of azithromycin resistance, was higher among MSM who had a BEP detected compared to those who did not (41.3% vs 14.1%). Furthermore, among people who had a BEP detected, *mphA* was associated with a previous bacterial STI diagnosis, suggesting that antimicrobial exposure for STI treatment acts as a selective pressure on a range of gut organisms. I used a diagnosis of a bacterial STI as a marker for previous antimicrobial exposure, since antimicrobial treatment was not available in the routine data sources utilised in this study. This is clearly an imperfect and non-specific measure of antimicrobial exposure leading to some mis-classification but did provide an objective measurement for an important variable. In Chapters 4 and 5, *S. flexneri* cases

belonging to sexual networks of MSM were more likely to have isolates that were highly-resistant to azithromycin compared to cases whose isolates were located within travel-associated lineages. 88.4% of confirmed MSM had isolates that harboured genotypic markers of azithromycin resistance, and despite this, some (14.0%; 6/43) were prescribed azithromycin. The main limitation of the latter finding was that the questionnaire wording for antimicrobial usage (see Appendix 4.1) was not specific enough to rule out treatment for co-infections (e.g. bacterial STIs) and this should be modified going forward.

Current guidelines recommend avoiding antimicrobial treatment for self-limiting acute gastroenteritis, particularly when the aetiology is not known.[26,29,60,67] However, antimicrobial treatment may be required if a person is severely ill or at risk of complications due to immunosuppression or other co-morbidities. Due to increasing resistance, microbiological confirmation and antimicrobial susceptibility testing are essential to guide the appropriate choice of antimicrobial.[26] When an individual is severely ill, a clinician might make the decision that empirical treatment is necessary. However, this should be discussed with a clinical microbiologist, including the potential for whether the case is likely to be MSM or not (given the association with azithromycin resistance), and guided by local susceptibility data to ensure the most effective antimicrobial is used.

Azithromycin has been used for the treatment of several bacterial STIs, for which the development and spread of resistance is also a growing problem. Sustained transmission of *Neisserria gonorrhoeae* exhibiting high-level resistance to azithromycin was reported across the UK from 2014 to 2017.[332] As a result, treatment guidelines were updated in early 2019 to remove azithromycin from the previously recommended first-line dual therapy for gonorrhoea (ceftriaxone and azithromycin).[333] An additional reason for removing azithromycin was the concern that it could lead to the development and spread of macrolide resistance in other non-target organisms.[333] This is of relevance to *Mycoplasma genitalium*, which has received wide-spread media coverage since 2018 for fears that it may be the next resistant superbug, given increasing resistance to azithromycin.[334,335]

Antimicrobial stewardship in MSM is a significant issue that is not currently given enough attention. The consequences of azithromycin exposure for STI treatment in MSM are evident

for bacterial STIs and BEPs. Further concerns about the overuse of antimicrobials in this population more generally have been raised given recent studies exploring the role of antimicrobial pre- or post-exposure prophylaxis for bacterial STIs.[336-338] Current evidence from two randomised controlled trials suggests that doxycycline prophylaxis may reduce the incidence of bacterial STIs in MSM.[337,338] However, use of STI prophylactic antimicrobials in the MSM population requires careful consideration before widespread adoption, particularly given the potential to select for resistance in STI pathogens and other non-target organisms. There is clearly a need to review current guidelines on antimicrobial use in MSM; the parallel syndemics of increasingly resistant sexually transmissible pathogens need to be addressed holistically.

6.2.4 Detection and surveillance of enteric pathogens in MSM

Whole Genome Sequencing (WGS) has now largely replaced other typing methods for the public health surveillance of BEPs and for outbreak detection. In Chapter 4, I showed that WGS data, when combined with epidemiological data from case questionnaires, provide unique insight into the likely route of infection, which might be used to better inform targeted prevention advice and contact tracing. This could help to ensure that people receive appropriate advice on preventing onward transmission or re-infection through subsequent sexual contact and are referred for further STI testing if indicated.

At a population level, my analysis demonstrates how WGS and epidemiological data can be used to understand how much transmission occurs through sexual contact. I have also shown that WGS data can reveal the importation of novel strains into sexual networks of MSM, through travel and/or non-sexual transmission in the community, and the spread and persistence of specific strains within and across different transmission networks which might enable the deployment of appropriate and specific public health measures. WGS also provides a mechanism by which to monitor trends in antimicrobial resistance over time through the detection of genotypic markers of resistance. Although the WGS analyses presented in this thesis have primarily focused on identifying and characterising *S. flexneri,* the approach might also be applied to other enteric pathogens that are routinely sequenced at PHE, particularly those that show signals of suggestive MSM transmission (e.g. a high M:F ratio or proportion

of adult men, and/or geographical trends) or are circulating among MSM attending SHCs (Chapter 3).

My analyses have shown the value of including questions on sexual identity and behaviour on structured HPT exposure questionnaires. Sexual identity and sexual behaviour were well completed at 88.9% and 92.5%, respectively, which demonstrates that most people are willing to answer these questions (Chapters 4 and 5). Based on my analyses, routine inclusion of these questions has since been implemented at a national level for shigellosis. However, the inclusion of questions on sexual identity and behaviour should be implemented for all enteric pathogens where public health routine is follow-up to help establish potential exposure through sexual transmission. I recognise that HPTs and environmental health colleagues may lack the necessary training or skills to ask appropriate sensitive questions about sexual identity or potentially stigmatising behaviours and to offer appropriate advice, or have concerns about doing so. To address these issues, PHE has developed training sessions to improve skills in sexual history taking.[339]

Among *S. flexneri* cases that were reported to the GBRU from the HPTs participating in the pilot, 39% had linked questionnaire data (Chapter 4), suggesting questionnaire completion is not always feasible, practical or timely enough. PHE is currently in the process of developing patient-facing online questionnaires, which may improve questionnaire completion in regions which have inadequate staffing capacity to undertake follow-up questionnaires for all cases. Use of an online questionnaire may also encourage patients to disclose stigmatising behaviours that they may not otherwise do when speaking directly to a healthcare professional.[340]

When the collection of detailed exposure data is neither feasible or timely enough, Single Nucleotide Polymorphism (SNP) typing can be used to identify transmission clusters in near real-time using only basic demographic data submitted alongside laboratory isolates. In Chapter 5, I validated the SNP clustering tool for identifying and distinguishing sexual and non-sexual transmission clusters of *S. flexneri* and this approach can be used by PHE to inform the public health response, even when data on sexual identity and behaviour are not available. Not only does this provide a mechanism for delivering a more rapid and appropriate public

health response, the tool provides opportunities for evaluating the effectiveness of control measures by monitoring long term trends including the spread of infection within and between MSM and non-MSM populations. In addition, it provides a potential mechanism for providing preventive advice to cases if shared in a sensitive manner which builds trust between the healthcare professional and the patient without deductive disclosure.[341] In addition to *Shigella* spp., SNP typing is routinely performed by PHE for a range of BEPs including pathogenic *E. coli* and *Salmonella* spp. and the utility of this tool for identifying and distinguishing sexual transmission will require validation on a pathogen-by-pathogen basis.

Globally, there has been a shift towards WGS for BEP surveillance and outbreak detection. In 2013, the National Enteric Reference Laboratory at the Centers for Disease Control and Prevention (CDC) in the USA began its steady transition towards WGS for BEP surveillance and outbreak detection and has since expanded this technology across the national public health laboratory network, known as PulseNet.[342,343] The CDC also manages PulseNet International, which is a global laboratory network for BEP surveillance that is working towards routine implementation of WGS using standardised protocols.[344] Of course, the ability to do this depends on local infrastructure and expertise. My analyses will be of value to other countries that are developing or currently using WGS for surveillance, particularly where BEP outbreaks have been reported among MSM. Laboratory networks, such as PulseNet International, allow WGS data to be shared across different countries to identify international sexual networks, monitor patterns of AMR, and enable collaborations that aim to control the spread of BEPs.

6.3 Further research

There are a number of areas of research which could be explored in the future to expand our understanding of the epidemiology of enteric pathogens in MSM to inform decision-making about control and the design of effective interventions.

i) Understanding the burden and risk factors of enteric pathogens in different MSM populations

To enable deeper understanding of the risk factors for BEPs in MSM, future studies could include different MSM populations, a larger sample of MSM living with HIV, and more detailed information on specific sexual practices (e.g. rimming) and symptom presentation. These data might be best collected using a bespoke questionnaire to allow greater exploration of specific characteristics and behaviours that may be relevant to the research questions. MSM attending a SHC outside of London or recruited through community venues and/or social media applications are potentially suitable study populations.

ii) Parameters that influence the spread of enteric pathogens

Further exploration of the duration of infectiousness as well as the risk of onward transmission from an asymptomatic carrier could inform the development of interventions and be explored through a prospective cohort study of MSM attending SHCs. The collection of faecal samples at regular intervals would provide detailed information on bacterial load, shedding time and clearance, and enable the distinction between re-infection or long-term carriage. Quantifying the risk of transmission directly is more challenging since this would require knowledge of sexual contacts and the direction of transmission.[221] Interpretation is further complicated by the type of sexual contact and immune status. Nonetheless, it would be possible to collect detailed information on the time and frequency of specific sexual practices and clinical symptoms (including whether sexual partners were symptomatic) with the use of daily or weekly diaries.

iii) Mathematical modelling to estimate the impact of interventions

Mathematical modelling could be used to better understand the transmission dynamics of BEPs in MSM and to estimate the potential impact of public health interventions and control strategies. Data generated from my research could be used to help parameterise the models including estimates of BEP prevalence, information about asymptomatic carriage, the number of new sexual partners reported, and patterns of mixing between MSM and non-MSM populations.

iv) Use of WGS data to identify and characterise sexual transmission of other BEPs

The routine implementation of WGS at PHE provides an invaluable opportunity to explore the transmission dynamics and genotypic markers of AMR in BEPs other than *Shigella* spp. with the aim of identifying and characterising potential sexual transmission between men. Expanding public health follow-up questionnaires to collect data on sexual identity and behaviour for other enteric pathogens is recommended and can complement the WGS data. At a population level, these data might help us to better understand how much transmission occurs through sexual contact, to explore the characteristics associated with different transmission networks and to enable the targeting of appropriate and specific public health measures.

v) The role of the gut microbiota

Further studies are required to determine whether gut microbiota patterns and prevalence of AMR genes differ between MSM and other adults. One approach would be to perform metagenomic sequencing and analysis on collected stool samples, which could be linked with data on STI diagnoses, antimicrobial use, sexual behaviour and HIV status. This could provide novel insights into the potential role of the gut microbiota in the development and severity of gastrointestinal disease and the transmission of AMR genes (i.e. does the gut microbiota act as a reservoir).

vi) Antimicrobial use and AMR

The development and spread of AMR in bacterial pathogens associated with MSM populations is a growing concern and further research is needed to understand the consequences of antimicrobial exposure on levels of AMR in BEPs at an individual and population level. Research is underway combining laboratory experimentation with mathematical modelling to understand the emergence and spread of AMR across different bacterial species and how this might contribute to epidemics in MSM (KS Baker, personal communication, June 2019). In addition, qualitative research exploring knowledge, perceptions and attitudes about antimicrobial use and AMR among MSM and healthcare providers could inform future guidelines, by improving understanding of prescribing practices and potential barriers to antimicrobial stewardship.

vii) Expand research to include protozoan and viral enteric pathogens

There has been limited research on the prevalence and risk factors for protozoan and viral enteric pathogens including *G. lamblia*, *E. histolytica* and HAV in MSM. Future studies could explore whether risk factors for these pathogens are similar to those of BEPs, since my literature review showed that they have been responsible for recent outbreaks in MSM.

6.4 Overall conclusions

My research has improved understanding of the epidemiology and transmission of enteric pathogens in MSM. I have demonstrated that a range of BEPs are circulating among MSM attending the UK's largest SHC and these are associated with higher-risk sexual behaviours in dense sexual networks that are also characterised by STIs, HIV and high levels of AMR. Asymptomatic carriage in this population may be hindering effective control. Those at higher risk of acquiring enteric pathogens could be targeted with specific health education messages to raise awareness, and sexual partners of those diagnosed with an enteric pathogen should be notified and given appropriate advice. However, controlling transmission is challenging and will require improved understanding of the social and psychological motivations for engaging in higher-risk behaviours, as well as a holistic approach to prevention and management that also considers the long-term consequences of exposure to antimicrobials. WGS provides a potential mechanism for identifying sexual transmission in MSM to ensure that appropriate interventions can be delivered rapidly.

References

1. Kotloff KL, Winickoff JP, Ivanoff B, et al. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ.* 1999;77(8):651-666.
2. Global Burden of Disease (GBD) Diarrhoeal Diseases Collaborators. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis.* 2017;17(9):909-948.
3. Kazal HL, Sohn N, Carrasco JI, Robilotti JG, Delaney WE. The gay bowel syndrome: clinico pathologic correlation in 260 cases. *Annals of Clinical and Laboratory Science.* 1976;6(2):184-192.
4. Gilbart VL, Simms I, Jenkins C, et al. Sex, drugs and smart phone applications: findings from semistructured interviews with men who have sex with men diagnosed with *Shigella flexneri* 3a in England and Wales. *Sexually Transmitted Infections.* 2015;91(8):598-602.
5. Simms I, Field N, Jenkins C, et al. Intensified shigellosis epidemic associated with sexual transmission in men who have sex with men - *Shigella flexneri* and *S. sonnei* in England, 2004 to end of February 2015. *Eurosurveillance.* 2015;20(15):2-6.
6. O'Sullivan B, Delpech V, Pontivivo G, et al. Shigellosis linked to sex venues, Australia. *Emerg Infect Dis.* 2002;8(8):862-864.
7. Marcus U, Zucs P, Bremer V, et al. Shigellosis - a re-emerging sexually transmitted infection: outbreak in men having sex with men in Berlin. *Int J STD AIDS.* 2004;15(8):533-537.
8. Beebeejaun K, Degala S, Balogun K, et al. Outbreak of hepatitis A associated with men who have sex with men (MSM), England, July 2016 to January 2017. *Eurosurveillance.* 2017;22(5):7-12.
9. Freidl GS, Sonder GJ, Bovee LP, et al. Hepatitis A outbreak among men who have sex with men (MSM) predominantly linked with the EuroPride, the Netherlands, July 2016 to February 2017. *Eurosurveillance.* 2017;22(8):2-6.
10. Werber D, Michaelis K, Hausner M, et al. Ongoing outbreaks of hepatitis A among men who have sex with men (MSM), Berlin, November 2016 to January 2017 – Linked to other German cities and European countries. *Eurosurveillance.* 2017;22(5).
11. Gaudreau C, Rodrigues-Coutlée S, Pilon PA, Coutlée F, Bekal S. Long-Lasting Outbreak of Erythromycin- and Ciprofloxacin-Resistant *Campylobacter jejuni*

Subspecies jejuni from 2003 to 2013 in Men Who Have Sex with Men, Quebec, Canada. *Clinical Infectious Diseases.* 2015;61(10):1549-1552.

12. Gaudreau C, Pilon PA, Sylvestre JL, Boucher F, Bekal S. Multidrug-Resistant *Campylobacter coli* in Men Who Have Sex with Men, Quebec, Canada, 2015. *Emerging Infectious Diseases.* 2016;22(9):1661-1663.

13. Simms I, Gilbart VL, Byrne L, et al. Identification of verocytotoxin-producing *Escherichia coli* O117:H7 in men who have sex with men, England, November 2013 to August 2014. *Eurosurveillance.* 2014;19(43):10-12.

14. Escola-Verge L, Arando M, Vall M, et al. Outbreak of intestinal amoebiasis among men who have sex with men, Barcelona (Spain), October 2016 and January 2017. *Euro Surveill.* 2017;22(30):27.

15. Borg ML, Modi A, Tostmann A, et al. Ongoing outbreak of *Shigella flexneri* serotype 3a in men who have sex with men in England and Wales, data from 2009-2011. *Eurosurveillance.* 2012;17(13):2-6.

16. European Centre for Disease Prevention and Control (ECDC). *Hepatitis A outbreak in the EU/EEA mostly affecting men who have sex with men -third update, 28 June 2017.* Stockholm: ECDC. 2017.

17. Gozlan Y, Bar-Or I, Rakovsky A, et al. Ongoing hepatitis A among men who have sex with men (MSM) linked to outbreaks in Europe in Tel Aviv area, Israel, December 2016 – June 2017. *Eurosurveillance.* 2017;22(29).

18. Rivas V, Barrera A, Pino K, et al. Hepatitis A outbreak since November 2016 affecting men who have sex with men (MSM) in Chile connected to the current outbreak in MSM in Europe, situation up to October 2017. *Euro Surveill.* 2018;23(9).

19. Chen GJ, Lin KY, Hung CC, Chang SC. Hepatitis A Outbreak Among Men Who Have Sex With Men in a Country of Low Endemicity of Hepatitis A Infection. *Journal of Infectious Diseases.* 2017;215(8):1339-1340.

20. Aghaizu A, Wayal S, Nardone A, et al. Sexual behaviours, HIV testing, and the proportion of men at risk of transmitting and acquiring HIV in London, UK, 2000-13: a serial cross-sectional study. *Lancet Hiv.* 2016;3(9):E431-E440.

21. Mohammed H, Mitchell H, Sile B, Duffell S, Nardone A, Hughes G. Increase in Sexually Transmitted Infections among Men Who Have Sex with Men, England, 2014. *Emerging Infectious Diseases.* 2016;22(1):88-91.

22. Baker KS, Dallman TJ, Ashton PM, et al. Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: A cross-sectional study. *The Lancet Infectious Diseases.* 2015;15(8):913-921.

23. Mook P, McCormick J, Bains M, et al. ESBL-Producing and Macrolide-Resistant *Shigella sonnei* Infections among Men Who Have Sex with Men, England, 2015. *Emerging Infectious Diseases.* 2016;22(11):1948-1952.

24. Baker KS, Dallman TJ, Field N, et al. Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. *Nat Commun.* 2018;9(1):1462.

25. The National Archives. The Health Protection (Notification) Regulations 2010. Available from: http://www.legislation.gov.uk/uksi/2010/659/contents/made [Accessed 20th April 2018]. Contains public sector information licensed under the Open Government Licence v3.0.

26. Clutterbuck D, Asboe D, Barber T, et al. 2016 United Kingdom national guideline on the sexual health care of men who have sex with men. *Int J STD AIDS.* 2018:956462417746897.

27. Public Health England. *Interim Public Health Operational Guidelines for Shigellosis: A joint guideline from Public Health England and the Chartered Institute of Environmental Health*. London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0

28. Hughes G, Silalang P, Were J, et al. Prevalence and characteristics of gastrointestinal infections in men who have sex with men diagnosed with rectal chlamydia infection in the UK: an 'unlinked anonymous' cross-sectional study. *Sex Transm Infect.* 2017;Published Online First: 30 March 2017.

29. Farthing M, Feldman R, Finch R, et al. The Management of Infective Gastroenteritis in Adults - A consensus statement by an expert panel convened by the British Society for the Study of Infection. *Journal of Infection.* 1996;33(3):143-152.

30. Farthing M, Salam M, Lindberg G, et al. *Acute diarrhea in adults and children: a global perspective - World Gastroenterology Organisation Global Guidelines*. USA: World Gastroenterology Organisation. 2012.

31. DuPont HL. Acute Infectious Diarrhea in Immunocompetent Adults. *New Engl J Med.* 2014;370(16):1532-1540.

32. Tam C, Viviani L, Adak B, et al. *The Second Study of Infectious Intestinal Disease in the Community (IID2 Study) Final report*. London: Food Standards Agency. 2012. Contains public sector information licensed under the Open Government Licence v3.0.

33. Quilliam RS, Cross P, Williams AP, et al. Subclinical infection and asymptomatic carriage of gastrointestinal zoonoses: occupational exposure, environmental pathways, and the anonymous spread of disease. *Epidemiol Infect.* 2013;141(10):2011-2021.

34. Tompkins DS, Hudson MJ, Smith HR, et al. A study of infectious intestinal disease in England: microbiological findings in cases and controls. *Commun Dis Public Health.* 1999;2(2):108-113.

35. Gillespie IA, O'Brien SJ, Adak GK, et al. Point source outbreaks of *Campylobacter jejuni* infection--are they more common than we think and what might cause them? *Epidemiol Infect.* 2003;130(3):367-375.

36. The former Public Health Laboratory Service (PHLS) Advisory Committee on Gastrointestinal Infections. Preventing person-to-person spread following gastrointestinal infections: guidelines for public health physicians and environmental health officers. *Commun Dis Public Health.* 2004;7(4):362-384.

37. O'Brien SJ, Halder SL. GI Epidemiology: infection epidemiology and acute gastrointestinal infections. *Aliment Pharmacol Ther.* 2007;25(6):669-674.

38. DuPont HL, Levine MM, Hornick RB, Formal SB. Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis.* 1989;159(6):1126-1128.

39. Public Health England. *Investigations of Faecal Specimens for Enteric Pathogens*: *UK Standards for Microbiology Investigations (B30 Issue 8.1)*. London: Public Health England. 2014. Contains public sector information licensed under the Open Government Licence v3.0.

40. Lund BM, O'Brien SJ. The occurrence and prevention of foodborne disease in vulnerable people. *Foodborne Pathog Dis.* 2011;8(9):961-973.

41. Nygard K, Andersson Y, Rottingen JA, et al. Association between environmental risk factors and campylobacter infections in Sweden. *Epidemiol Infect.* 2004;132(2):317-325.

42. Byrne L, Jenkins C, Launders N, Elson R, Adak GK. The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol Infect.* 2015;143(16):3475-3487.

43. World Health Organization. *Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1*. Geneva: World Health Organization. 2005.

44. Livio S, Strockbine NA, Panchalingam S, et al. *Shigella* isolates from the global enteric multicenter study inform vaccine development. *Clin Infect Dis.* 2014;59(7):933-941.

45. Dallman TJ, Chattaway MA, Mook P, Godbole G, Crook PD, Jenkins C. Use of whole-genome sequencing for the public health surveillance of *Shigella sonnei* in England and wales, 2015. *Journal of Medical Microbiology.* 2016;65(8):882-884.

46. Public Health England and the Chartered Institute of Environmental Health. *Recommendations for the Public Health Management of Gastrointestinal Infections*

2019: Principles and practice. London: Public Health England. 2020. Contains public sector information licensed under the Open Government Licence v3.0.

47. Noriega FR, Liao FM, Formal SB, Fasano A, Levine MM. Prevalence of *Shigella* enterotoxin 1 among *Shigella* clinical isolates of diverse serotypes. *J Infect Dis.* 1995;172(5):1408-1410.

48. Hawker J, Begg N, Blair I, Reintjes R, Weinberg J, Ekdahl K. *Communicable Disease Control and Health Protection Handbook (Third edition).* Oxford: Blackwell Publishing; 2012. Copyright 2012 Jeremy Hawker, Norman Begg, Iain Blair, Ralf Reintjes, Julius Weinberg and Karl Ekdahl. Published 2012 by Blackwell Publishing Ltd.

49. Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. Experimental *Campylobacter jejuni* infection in humans. *J Infect Dis.* 1988;157(3):472-479.

50. Janssen R, Krogfelt KA, Cawthraw SA, van Pelt W, Wagenaar JA, Owen RJ. Host-pathogen interactions in *Campylobacter* infections: the host perspective. *Clin Microbiol Rev.* 2008;21(3):505-518.

51. Esan OB, Pearce M, van Hecke O, et al. Factors Associated with Sequelae of *Campylobacter* and Non-typhoidal *Salmonella* Infections: A Systematic Review. *EBioMedicine.* 2017;15:100-111.

52. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet.* 2005;365(9464):1073-1086.

53. Haque R, Huston CD, Hughes M, Houpt E, Petri WA, Jr. Amebiasis. *N Engl J Med.* 2003;348(16):1565-1573.

54. Public Health England. *Gastroenteritis and Diarrhoea: UK Standards for Microbiology Investigations (S7 Issue 1)*. London: Public Health England. 2013. Contains public sector information licensed under the Open Government Licence v3.0.

55. Public Health England. *Investigation of specimens other than blood for parasites: UK Standards for Microbiology Investigations (B31 Issue 5.1)*. London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0.

56. Chalmers RM, Campbell BM, Crouch N, Charlett A, Davies AP. Comparison of diagnostic sensitivity and specificity of seven *Cryptosporidium* assays used in the UK. *J Med Microbiol.* 2011;60(Pt 11):1598-1604.

57. Alexander CL, Currie S, Pollock K, Smith-Palmer A, Jones BL. An audit of *Cryptosporidium* and *Giardia* detection in Scottish National Health Service Diagnostic Microbiology Laboratories. *Epidemiol Infect.* 2017;145(8):1584-1590.

58. Public Health England. *Hepatitis A Virus Acute Infection Serology: UK Standards for Microbiology Investigations (V27 Issue 3)*. London: Public Health England. 2014. Contains public sector information licensed under the Open Government Licence v3.0.

59. Public Health England. *National Infection Service Reference Laboratories Colindale: Bacteriology Reference Department User Manual (Version 9).* London: Public Health England. 2018. Contains public sector information licensed under the Open Government Licence v3.0.

60. Public Health England. *Managing suspected infectious diarrhoea: Quick reference guidance for primary care*. London: Public Health England. 2015. Contains public sector information licensed under the Open Government Licence v3.0.

61. De Bruyn G, Hahn S, Borwick A. Antibiotic treatment for travellers' diarrhoea. *Cochrane Database Syst Rev.* 2000(3):CD002242.

62. Wistrom J, Gentry LO, Palmgren AC, et al. Ecological effects of short-term ciprofloxacin treatment of travellers' diarrhoea. *J Antimicrob Chemother.* 1992;30(5):693-706.

63. Neill MA, Opal SM, Heelan J, et al. Failure of ciprofloxacin to eradicate convalescent fecal excretion after acute salmonellosis: experience during an outbreak in health care workers. *Ann Intern Med.* 1991;114(3):195-199.

64. Wong CS, Jelacic S, Habeeb RL, Watkins SL, Tarr PI. The risk of the hemolytic-uremic syndrome after antibiotic treatment of Escherichia coli O157:H7 infections. *N Engl J Med.* 2000;342(26):1930-1936.

65. Wong CS, Mooney JC, Brandt JR, et al. Risk factors for the hemolytic uremic syndrome in children infected with *Escherichia coli* O157:H7: a multivariable analysis. *Clin Infect Dis.* 2012;55(1):33-41.

66. Freedman SB, Xie J, Neufeld MS, et al. Shiga Toxin-Producing *Escherichia coli* Infection, Antibiotics, and Risk of Developing Hemolytic Uremic Syndrome: A Meta-analysis. *Clin Infect Dis.* 2016;62(10):1251-1258.

67. The National Institute for Health and Care Excellence (NICE). Gastroenteritis. Availbale from: https://cks.nice.org.uk/gastroenteritis [Accessed 22nd October 2019].

68. BNF. Antiprotozoal drugs. Available from: https://bnf.nice.org.uk/treatment-summary/antiprotozoal-drugs.html [Accessed 18th October 2018].

69. M'ikanatha NM, Lynfield R, Julian KG, Van Beneden CA, Valk Hd. Infectious disease surveillance: a cornerstone for prevention and control. In: M'ikanatha NM, Julian KG, Van Beneden CA, Valk Hd, eds. *Infectious Disease Surveillance, Second Edition.* Oxford: Wiley & Sons Ltd; 2013:3-20.

70. Public Health England. *PHE Gastrointestinal Infections: Data Summary of eFOSS data, 2013*. London: Public Health England. 2013. Contains public sector information licensed under the Open Government Licence v3.0.

71. Public Health England. Gastrointestinal infections: guidance, data and analysis. https://www.gov.uk/government/collections/gastrointestinal-infections-guidance-data-and-analysis. [Accessed 29th March 2017].

72. Department of Health. *Health Protection Legislation (England) Guidance 2010.* London: Department of Health. 2010. Contains public sector information licensed under the Open Government Licence v3.0.

73. Public Health England. *Laboratory reporting to Public Health England: A guide for diagnostic laboratories*. London: Public Health England. 2016. Contains public sector information licensed under the Open Government Licence v3.0.

74. O'Brien SJ, Rait G, Hunter PR, et al. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). *BMC Med Res Methodol.* 2010;10:39.

75. Lecky DM, Hawking MK, McNulty CA, group Es. Patients' perspectives on providing a stool sample to their GP: a qualitative study. *Br J Gen Pract.* 2014;64(628):e684-693.

76. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut.* 2012;61(1):69-77.

77. Sethi D, Wheeler JG, Cowden JM, et al. A study of infectious intestinal disease in England: plan and methods of data collection. *Commun Dis Public Health.* 1999;2(2):101-107.

78. Wheeler JG, Sethi D, Cowden JM, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *BMJ.* 1999;318(7190):1046-1050.

79. Food Standards Agency. *A Report of the Study of Infectious Intestinal Disease in England*. London: Food Standards Agency. 2000. Contains public sector information licensed under the Open Government Licence v3.0.

80. Tam CC, O'Brien SJ, Tompkins DS, et al. Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. *Clin Infect Dis.* 2012;54(9):1275-1286.

81. Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). *Eur J Clin Microbiol Infect Dis.* 2007;26(5):311-323.

82. Public Health England. *Interim Public Health Operational Guidedance for Shiga toxin producing Escherichia coli (STEC) including STEC (0157 and non-0157) infections*. London: Public Health England. 2018. Contains public sector information licensed under the Open Government Licence v3.0.

83. Public Health England. *Communicable Disease Outbreak Management: Operational guidance*. London: Public Health England. 2014. Contains public sector information licensed under the Open Government Licence v3.0.

84. Public Health England. *Public health control and management of hepatitis A: 2017 Guidelines*. London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0.

85. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc.* 2015;13(3):141-146.

86. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Social Research Methodology.* 2005;8(1):19-32.

87. Weller IVD. The gay bowel. *Gut.* 1985;26(9):869-875.

88. Dritz SK, Ainsworth TE, Back A, et al. Patterns of sexually transmitted enteric diseases in a city. *Lancet.* 1977;2(8027):3-4.

89. Kaufman JC, Fierst SM. Shigellosis and the gay bowel syndrome: an endoscopic point of view and review of the literature. *Gastrointestinal Endoscopy.* 1982;28(4):250-251.

90. Bader M, Pedersen AH, Williams R, Spearman J, Anderson H. Venereal transmission of shigellosis in Seattle-King county. *Sexually Transmitted Diseases.* 1977;4(3):89-91.

91. Dritz SK, Braff EH. Sexually transmitted typhoid fever. *N Engl J Med.* 1977;296(23):1359-1360.

92. Meyers JD, Kuharic HA, Holmes KK. *Giardia lamblia* infection in homosexual men. *Br J Vener Dis.* 1977;53(1):54-55.

93. Keystone JS, Keystone DL, Proctor EM. Intestinal parasitic infections in homosexual men: prevalence, symptoms and factors in transmission. *Can Med Assoc J.* 1980;123(6):512-514.

94. Phillips SC, Mildvan D, William DC, Gelb AM, White MC. Sexual transmission of enteric protozoa and helminths in a venereal-disease-clinic population. *N Engl J Med.* 1981;305(11):603-606.

95. Markell EK, Havens RF, Kuritsubo RA, Wingerd J. Intestinal protozoa in homosexual men of the San Francisco Bay area: prevalence and correlates of infection. *Am J Trop Med Hyg.* 1984;33(2):239-245.

96. Hughes G, Field N. The epidemiology of sexually transmitted infections in the UK: impact of behavior, services and interventions. *Future Microbiol.* 2015;10(1):35-51.

97. Nicoll A, Hughes G, Donnelly M, et al. Assessing the impact of national anti-HIV sexual health campaigns: trends in the transmission of HIV and other sexually transmitted infections in England. *Sexually Transmitted Infections.* 2001;77(4):242-247.

98. Nicoll A, Hamers FF. Are trends in HIV, gonorrhoea, and syphilis worsening in western Europe? *BMJ.* 2002;324(7349):1324-1327.

99. Elford J. Changing patterns of sexual behaviour in the era of highly active antiretroviral therapy. *Curr Opin Infect Dis.* 2006;19(1):26-32.

100. Strauss B, Kurzac C, Embree G, Sevigny R, Paccagnella A, Fyfe M. Clusters of *Shigella sonnei* in men who have sex with men, British Columbia, 2001. *Can Commun Dis Rep.* 2001;27(13):109-110; discussion 110-104.

101. Centers for Disease Control and Prevention. Increases in unsafe sex and rectal gonorrhea among men who have sex with men--San Francisco, California, 1994-1997. *MMWR Morb Mortal Wkly Rep.* 1999;48(3):45-48.

102. Centers for Disease Control and Prevention. *Shigella sonnei* outbreak among men who have sex with men--San Francisco, California, 2000-2001. *MMWR Morb Mortal Wkly Rep.* 2001;50(42):922-926.

103. de Vries HJ, Zingoni A, White JA, Ross JD, Kreuter A. 2013 European Guideline on the management of proctitis, proctocolitis and enteritis caused by sexually transmissible pathogens. *Int J STD AIDS.* 2014;25(7):465-474.

104. Quinn TC. Sexually transmitted enteric infections. In: Schlossberg D, ed. *Clinical Infectious Disease. .* Second Edition ed. Cambridge: Cambridge University Press; 2015:352-356.

105. Leentvaar-Kuijpers A, Kool JL, Veugelers PJ, Coutinho RA, van Griensven GJ. An outbreak of hepatitis A among homosexual men in Amsterdam, 1991-1993. *Int J Epidemiol.* 1995;24(1):218-222.

106. Reintjes R, Bosman A, de Zwart O, Stevens M, van der Knaap L, van den Hoek K. Outbreak of hepatitis A in Rotterdam associated with visits to 'darkrooms' in gay bars. *Commun Dis Public Health.* 1999;2(1):43-46.

107. Mckenna KYA, Green AS, Smith PK. Demarginalizing the sexual self. *The Journal of Sex Research.* 2001;38(4):302-311.

108. Henning KJ, Bell E, Braun J, Barker ND. A community-wide outbreak of hepatitis A: risk factors for infection among homosexual and bisexual men. *Am J Med.* 1995;99(2):132-136.

109. Bourne A, Reid D, Hickson F, Torres Rueda S, Weatherburn P. *The chemsex study: Drug use in sexual settings among gay and bisexual men in Lambeth, Southwark and Lewisham*. London: Sigma Research, London School of Hygiene and Tropical Medicine. 2014.

110. Theodore PS, Duran RE, Antoni MH. Drug use and sexual risk among gay and bisexual men who frequent party venues. *AIDS Behaviour.* 2014;18:2178-2186.

111. Mattison AM, Ross MW, Wolfson T, Franklin D, San Diego HIVNRCG. Circuit party attendance, club drug use, and unsafe sex in gay men. *J Subst Abuse.* 2001;13(1-2):119-126.

112. Sewell J, Miltz A, Lampe FC, et al. Poly drug use, chemsex drug use, and associations with sexual risk behaviour in HIV-negative men who have sex with men attending sexual health clinics. *Int J Drug Policy.* 2017;43:33-43.

113. Forrest DW, Metsch LR, LaLota M, Cardenas G, Beck DW, Jeanty Y. Crystal methamphetamine use and sexual risk behaviors among HIV-positive and HIV-negative men who have sex with men in South Florida. *J Urban Health.* 2010;87(3):480-485.

114. Daskalopoulou M, Rodger A, Phillips AN, et al. Recreational drug use, polydrug use, and sexual behaviour in HIV-diagnosed men who have sex with men in the UK: results from the cross-sectional ASTRA study. *Lancet HIV.* 2014;1(1):e22-31.

115. Weatherburn P, Hickson F, Reid D, Torres-Rueda S, Bourne A. Motivations and values associated with combining sex and illicit drugs ('chemsex') among gay men in South London: findings from a qualitative study. *Sex Transm Infect.* 2017;93(3):203-206.

116. Ahmed AK, Weatherburn P, Reid D, et al. Social norms related to combining drugs and sex ("chemsex") among gay men in South London. *Int J Drug Policy.* 2016;38:29-35.

117. Hughes G, Alexander S, Simms I, et al. Lymphogranuloma venereum diagnoses among men who have sex with men in the U.K.: interpreting a cross-sectional study using an epidemic phase-specific framework. *Sex Transm Infect.* 2013;89(7):542-547.

118. Simms I, Wallace L, Thomas DR, et al. Recent outbreaks of infectious syphilis, United Kingdom, January 2012 to April 2014. *Euro Surveill.* 2014;19(24).

119. Hung CC, Chang SY, Ji DD. *Entamoeba histolytica* infection in men who have sex with men. *The Lancet Infectious Diseases.* 2012;12(9):729-736.

120. Baer JT, Vugia DJ, Reingold AL, Aragon T, Angulo FJ, Bradford WZ. HIV infection as a risk factor for shigellosis. *Emerg Infect Dis.* 1999;5(6):820-823.

121. Mohan K, Hibbert M, Rooney G, et al. What is the overlap between HIV and shigellosis epidemics in England: further evidence of MSM transmission? *Sexually Transmitted Infections.* 2017;10:10.

122. Aragon TJ, Vugia DJ, Shallow S, et al. Case-control study of shigellosis in San Francisco: the role of sexual transmission and HIV infection. *Clin Infect Dis.* 2007;44(3):327-334.

123. Daskalakis DC, Blaser MJ. Another perfect storm: *Shigella*, men who have sex with men, and HIV. *Clin Infect Dis.* 2007;44(3):335-337.

124. Bottone EJ, Wormser GP, Duncanson FP. Nontyphoidal *Salmonella* bacteremia as an early infection in acquired immunodeficiency syndrome. *Diagn Microbiol Infect Dis.* 1984;2(3):247-250.

125. Jacobs JL, Gold JW, Murray HW, Roberts RB, Armstrong D. Salmonella infections in patients with the acquired immunodeficiency syndrome. *Ann Intern Med.* 1985;102(2):186-188.

126. Mayer KH, Hanson E. Recurrent salmonella infection with a single strain in the acquired immunodeficiency syndrome. Confirmation by plasmid fingerprinting. *Diagn Microbiol Infect Dis.* 1986;4(1):71-76.

127. Simor AE, Poon R, Borczyk A. Chronic *Shigella flexneri* infection preceding development of acquired immunodeficiency syndrome. *J Clin Microbiol.* 1989;27(2):353-355.

128. Blaser MJ, Hale TL, Formal SB. Recurrent shigellosis complicating human immunodeficiency virus infection: failure of pre-existing antibodies to confer protection. *Am J Med.* 1989;86(1):105-107.

129. Mandell W, Neu H. *Shigella* bacteremia in adults. *JAMA.* 1986;255(22):3116-3117.

130. Nelson MR, Shanson DC, Hawkins DA, Gazzard BG. Salmonella, Campylobacter and Shigella in HIV-seropositive patients. *AIDS.* 1992;6(12):1495-1498.

131. Wilmer A, Romney MG, Gustafson R, et al. *Shigella flexneri* serotype 1 infections in men who have sex with men in Vancouver, Canada. *HIV Medicine.* 2015;16(3):168-175.

132. Cresswell FV, Ross S, Booth T, et al. *Shigella flexneri*: A Cause of Significant Morbidity and Associated With Sexually Transmitted Infections in Men Who Have Sex With Men. *Sexually Transmitted Diseases.* 2015;42(6):344-344.

133. Celum CL, Chaisson RE, Rutherford GW, Barnhart JL, Echenberg DF. Incidence of salmonellosis in patients with AIDS. *J Infect Dis.* 1987;156(6):998-1002.

134. Sorvillo FJ, Lieb LE, Waterman SH. Incidence of campylobacteriosis among patients with AIDS in Los Angeles County. *J Acquir Immune Defic Syndr.* 1991;4(6):598-602.

135. Angulo FJ, Swerdlow DL. Bacterial enteric infections in persons infected with human immunodeficiency virus. *Clin Infect Dis.* 1995;21 Suppl 1:S84-93.

136. Marchand-Senecal X, Bekal S, Pilon PA, Sylvestre JL, Gaudreau C. Campylobacter fetus Cluster Among Men Who Have Sex With Men, Montreal, Quebec, Canada, 2014-2016. *Clin Infect Dis.* 2017.

137. James R, Barratt J, Marriott D, Harkness J, Stark D. Seroprevalence of *Entamoeba histolytica* infection among men who have sex with men in Sydney, Australia. *Am J Trop Med Hyg.* 2010;83(4):914-916.

138. Public Health England. HIV: annual data tables: National HIV surveillance data tables. Available from: https://www.gov.uk/government/statistics/hiv-annual-data-tables [Accessed 28th December 2019].

139. Public Health England. *HIV in the UK - 2016 report.* London: Public Health England. 2016. Contains public sector information licensed under the Open Government Licence v3.0.

140. Public Health England. *HIV in the United Kingdom: Towards Zero 2030 - 2019 report.* London: Public Health England. 2019. Contains public sector information licensed under the Open Government Licence v3.0.

141. Rodger AJ, Cambiano V, Bruun T, et al. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet.* 2019;393(10189):2428-2438.

142. Prevention Access Campaign. Undetectable=Untransmittable. Available from: https://www.preventionaccess.org/undetectable [Accessed 20th November 2019].

143. Malek R, Mitchell H, Furegato M, et al. Contribution of transmission in HIV-positive men who have sex with men to evolving epidemics of sexually transmitted infections in England: an analysis using multiple data sources, 2009-2013. *Euro Surveill.* 2015;20(15).

144. Daskalopoulou M, Rodger AJ, Phillips AN, et al. Condomless sex in HIV-diagnosed men who have sex with men in the UK: prevalence, correlates, and implications for HIV transmission. *Sex Transm Infect.* 2017;93(8):590-598.

145. Liao YS, Liu YY, Lo YC, Chiou CS. Azithromycin-Nonsusceptible *Shigella flexneri* 3a in Men Who Have Sex with Men, Taiwan, 2015-2016. *Emerging Infectious Diseases.* 2017;23(2):345-346.

146. Okame M, Adachi E, Sato H, et al. *Shigella sonnei* Outbreak among Men Who Have Sex with Men in Tokyo. *Japanese Journal of Infectious Diseases.* 2012;65(3):277-278.

147. Ratnayake R, Allard R, Pilon PA. Shifting dominance of *Shigella* species in men who have sex with men. *Epidemiology and Infection.* 2012;140(11):2082-2086.

148. Morgan O, Crook P, Cheasty T, et al. *Shigella sonnei* outbreak among homosexual men, London. *Emerg Infect Dis.* 2006;12(9):1458-1460.

149. Health Protection Agency. Outbreak of UK acquired *Shigella flexneri* in men who have sex with men. *Health Protection Report.* 2011;5(40).

150. Childs T, Simms I, Alexander S, Eastick K, Hughes G, Field N. Rapid increase in lymphogranuloma venereum in men who have sex with men, United Kingdom, 2003 to September 2015. *Euro Surveill.* 2015;20(48):30076.

151. Bernstein K, Bowen VB, Kim CR, et al. Re-emerging and newly recognized sexually transmitted infections: Can prior experiences shed light on future identification and control? *PLoS Med.* 2017;14(12):e1002474.

152. Gupta A, Polyak CS, Bishop RD, Sobel J, Mintz ED. Laboratory-confirmed shigellosis in the United States, 1989-2002: epidemiologic trends and patterns. *Clin Infect Dis.* 2004;38(10):1372-1377.

153. Tauxe RV, McDonald RC, Hargrett-Bean N, Blake PA. The persistence of *Shigella flexneri* in the United States: increasing role of adult males. *Am J Public Health.* 1988;78(11):1432-1435.

154. Cohen D, Green MS, Block C, Slepon R, Ofek I. Prospective study of the association between serum antibodies to lipopolysaccharide O antigen and the attack rate of shigellosis. *J Clin Microbiol.* 1991;29(2):386-389.

155. Cohen D, Bassal R, Goren S, et al. Recent trends in the epidemiology of shigellosis in Israel. *Epidemiol Infect.* 2014;142(12):2583-2594.

156. Calderon-Margalit R, Navon-Venezia S, Gefen D, et al. Biennial hyperepidemic shigellosis in an observant Jewish community. *Epidemiol Infect.* 2010;138(2):244-252.

157. The London HIV Prevention Programme (LHPP). Do It London. Available from: http://doitlondon.org/. [Accessed 5th May 2018].

158. Public Health England. *Update on the epidemiology of Shigella in adults in London, 2012-2016.* London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0.

159. Lane CR, Sutton B, Valcanis M, et al. Travel Destinations and Sexual Behavior as Indicators of Antibiotic Resistant *Shigella* Strains-Victoria, Australia. *Clinical Infectious Diseases.* 2016;62(6):722-729.

160. Bowen A, Grass J, Bicknese A, Campbell D, Hurd J, Kirkcaldy RD. Elevated risk for antimicrobial drug-resistant shigella infection among men who have sex with men, United States, 2011-2015. *Emerging Infectious Diseases.* 2016;22(9):1613-1616.

161. Gu B, Cao Y, Pan S, et al. Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of *Shigella* between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents.* 2012;40(1):9-17.

162. World Health Organization. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. Available from: https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/ [Accessed 12th October 2019]. 2017.

163. Chiou CS, Izumiya H, Kawamura M, et al. The worldwide spread of ciprofloxacin-resistant *Shigella sonnei* among HIV-infected men who have sex with men, Taiwan. *Clinical Microbiology and Infection.* 2016;22(4):383.e311-383.e316.

164. Hoffmann C, Sahly H, Jessen A, et al. High rates of quinolone-resistant strains of Shigella sonnei in HIV-infected MSM. *Infection.* 2013;41(5):999-1003.

165. Valcanis M, Brown JD, Hazelton B, et al. Outbreak of locally acquired azithromycin-resistant *Shigella flexneri* infection in men who have sex with men. *Pathology.* 2015;47(1):87-88.

166. Bowen A, Eikmeier D, Talley P, et al. Outbreaks of *Shigella sonnei* infection with decreased susceptibility to azithromycin among men who have sex with men — Chicago and metropolitan minneapolis-St. Paul, 2014. *Morbidity and Mortality Weekly Report.* 2015;64(21):597-598.

167. Gaudreau C, Barkati S, Leduc JM, Pilon PA, Favreau J, Bekal S. *Shigella* spp. with Reduced Azithromycin Susceptibility, Quebec, Canada, 2012-2013. *Emerging Infectious Diseases.* 2014;20(5):854-856.

168. Baker KS, Dallman TJ, Field N, et al. Genomic epidemiology of *Shigella* in the United Kingdom shows transmission of pathogen sublineages and determinants of antimicrobial resistance. *Sci Rep.* 2018;8(1):7389.

169. Ingle DJ, Easton M, Valcanis M, et al. Co-circulation of multidrug-resistant *Shigella* among men who have sex with men, Australia. *Clin Infect Dis.* 2019.

170. Koff RS. Clinical manifestations and diagnosis of hepatitis A virus infection. *Vaccine.* 1992;10 Suppl 1:S15-17.

171. Stene-Johansen K, Tjon G, Schreier E, et al. Molecular epidemiological studies show that hepatitis A virus is endemic among active homosexual men in Europe. *Journal of Medical Virology.* 2007;79(4):356-365.

172. Public Health England. Immunisation against infectious disease, 'Green Book'. Chapter 17: Hepatitis A v2.0. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/263309/Green_Book_Chapter_17_v2_0.pdf [Accessed 12th April 2018].

173. European Centre for Disease Prevention and Control (ECDC). *Hepatitis A outbreaks in the EU/EEA mostly affecting men who have sex with men.* Stockholm: ECDC. 2016.

174. European Centre for Disease Prevention and Control (ECDC). *Epidemiological update: hepatitis A outbreak in the EU/EEA mostly affecting men who have sex with men.* Stockholm: ECDC. 2017.

175. Departamento de Epidemiologia Ministerio de Salud. Informe de situacion epidemiologica de Hepatitis A y viral sin especificacion. Available from: http://epi.minsal.cl/wp-content/uploads/2017/05/Informe_HepA_SE182017.pdf. [Accessed 12th September 2017]. 2017.

176. New York City Department of Health and Mental Hygiene (DOHMH). Health department releases health alert on increase in hepatitis A cases among unvaccinated men who have sex with men. Available from: https://www1.nyc.gov/site/doh/about/press/pr2017/pr007-17.page [Accessed 12th September 2017]. 2017.

177. New South Wales Government. Hepatitis A outbreak related to person-to-person transmission in Sydney 2017/18. Available from: https://www.health.nsw.gov.au/Infectious/alerts/Pages/hep-A-outbreak-person-to-person.aspx [Accessed 2nd September 2019].

178. Plunkett J, Mandal S, Balogun K, et al. Hepatitis A outbreak among men who have sex with men (MSM) in England, 2016-2018: The contribution of past and current vaccination policy and practice. *Vaccine X.* 2019;1:100014.

179. Ndumbi P, Freidl GS, Williams CJ, et al. Hepatitis A outbreak disproportionately affecting men who have sex with men (MSM) in the European Union and European Economic Area, June 2016 to May 2017. *Euro Surveill.* 2018;23(33).

180. Chen GJ, Lin KY, Sun HY, et al. Incidence of acute hepatitis A among HIV-positive patients during an outbreak among MSM in Taiwan: Impact of HAV vaccination. *Liver Int.* 2017;08:08.

181. Brook G, Bhagani S, Kulasegaram R, et al. United Kingdom National Guideline on the Management of the viral hepatitides A, B and C 2015. *Int J STD AIDS.* 2016;27(7):501-525.

182. Clinical Effectiveness Group British Association for Sexual Health and HIV. 2017 interim update of the 2015 BASHH National Guidelines for the Management of the Viral Hepatitides. Available from :https://www.bashhguidelines.org/media/1161/viral-hepatitides-2017-update-18-12-17.pdf [Accessed 20th April 2018]. 2017.

183. Regan DG, Wood JG, Benevent C, et al. Estimating the critical immunity threshold for preventing hepatitis A outbreaks in men who have sex with men. *Epidemiology and Infection.* 2016;144(7):1528-1537.

184. Stark D, van Hal SJ, Matthews G, Harkness J, Marriott D. Invasive amebiasis in men who have sex with men, Australia. *Emerg Infect Dis.* 2008;14(7):1141-1143.

185. Lo YC, Ji DD, Hung CC. Prevalent and Incident HIV Diagnoses among Entamoeba histolytica-Infected Adult Males: A Changing Epidemiology Associated with Sexual Transmission — Taiwan, 2006–2013. *PLoS Neglected Tropical Diseases.* 2014;8(10).

186. Ohnishi K, Kato Y, Imamura A, et al. Present characteristics of symptomatic *Entamoeba histolytica* infection in the big cities of Japan. *Epidemiol Infect.* 2004;132(1):57-60.

187. Park WB, Choe PG, Jo JH, et al. Amebic liver abscess in HIV-infected patients, Republic of Korea. *Emerg Infect Dis.* 2007;13(3):516-517.

188. Wu H, Wu PY, Li SY, et al. Maximising the potential of voluntary counselling and testing for HIV: sexually transmitted infections and HIV epidemiology in a population testing for HIV and its implications for practice. *Sex Transm Infect.* 2012;88(8):612-616.

189. Kaakoush NO, Castano-Rodriguez N, Mitchell HM, Man SM. Global Epidemiology of *Campylobacter* Infection. *Clin Microbiol Rev.* 2015;28(3):687-720.

190. Man SM. The clinical importance of emerging *Campylobacter* species. *Nat Rev Gastroenterol Hepatol.* 2011;8(12):669-685.

191. Tee W, Mijch A. *Campylobacter jejuni* bacteremia in human immunodeficiency virus (HIV)-infected and non-HIV-infected patients: comparison of clinical features and review. *Clin Infect Dis.* 1998;26(1):91-96.

192. Wagenaar JA, van Bergen MA, Blaser MJ, Tauxe RV, Newell DG, van Putten JP. *Campylobacter fetus* infections in humans: exposure and disease. *Clin Infect Dis.* 2014;58(11):1579-1586.

193. BNF. Gastro-intestinal system infections, antibacterial therapy. Available from: https://bnf.nice.org.uk/treatment-summary/gastro-intestinal-system-infections-antibacterial-therapy.html [Accessed 30th August 2019].

194. Gaudreau C, Michaud S. Cluster of erythromycin- and ciprofloxacin-resistant *Campylobacter jejuni* subsp. jejuni from 1999 to 2001 in men who have sex with men, Quebec, Canada. *Clin Infect Dis.* 2003;37(1):131-136.

195. Gaudreau C, Helferty M, Sylvestre JL, et al. *Campylobacter coli* outbreak in men who have sex with men, Quebec, Canada, 2010-2011. *Emerging Infectious Diseases.* 2013;19(5):764-767.

196. Tee W, Mijch A, Wright E, Yung A. Emergence of multidrug resistance in *Campylobacter jejuni* isolates from three patients infected with human immunodeficiency virus. *Clin Infect Dis.* 1995;21(3):634-638.

197. Cooper F, Barber T. 'Gay bowel syndrome': Relic or real (and returning) phenomenon? *Current Opinion in Infectious Diseases.* 2014;27(1):84-89.

198. Dallman T, Cross L, Bishop C, et al. Whole genome sequencing of an unusual serotype of Shiga toxin-producing *Escherichia coli. Emerg Infect Dis.* 2013;19(8):1302-1304.

199. Baker KS, Dallman TJ, Thomson NR, Jenkins C. An outbreak of a rare Shiga-toxin-producing *Escherichia coli* serotype (O117:H7) among men who have sex with men. *Microb Genom.* 2018;4(7).

200. Public Health England. Laboratory surveillance of non-travel associated *Shigella* spp. infection in adult males, England: 2004 to 2017. *Health Protection Report.* 2017;11(42). Contains public sector information licensed under the Open Government Licence v3.0.

201. Mook P, Gardiner D, Kanagarajah S, et al. Use of gender distribution in routine surveillance data to detect potential transmission of gastrointestinal infections among men who have sex with men in England. *Epidemiol Infect.* 2018;146(11):1468-1477.

202. Beltrami JF, Shouse RL, Blake PA. Trends in infectious diseases and the male to female ratio: possible clues to changes in behavior among men who have sex with men. *AIDS Educ Prev.* 2005;17(6 Suppl B):49-59.

203. Narayan S, Galanis E, Group BS. Are enteric infections sexually transmitted in British Columbia? *Can Commun Dis Rep.* 2016;42(2):24-29.

204. Syed NA, Hearing SD, Shaw IS, et al. Outbreak of hepatitis A in the injecting drug user and homeless populations in Bristol: control by a targeted vaccination programme and possible parenteral transmission. *Eur J Gastroenterol Hepatol.* 2003;15(8):901-906.

205. Hay G, Gannon M, MacDougall J, Eastwood C, Williams K, Millar T. Capture--recapture and anchored prevalence estimation of injecting drug users in England: national and regional estimates. *Stat Methods Med Res.* 2009;18(4):323-339.

206. Hines JZ, Pinsent T, Rees K, et al. Shigellosis outbreak among men who have sex with men and homeless persons — Oregon, 2015–2016. *Morbidity and Mortality Weekly Report.* 2016;65(31):812-813.

207. Ortega HB, Borchardt KA, Hamilton R, Ortega P, Mahood J. Enteric pathogenic protozoa in homosexual men from San Francisco. *Sex Transm Dis.* 1984;11(2):59-63.

208. Kean BH, William DC, Luminais SK. Epidemic of amoebiasis and giardiasis in a biased population. *Br J Vener Dis.* 1979;55(5):375-378.

209. William DC, Shookhoff HB, Felman YM, DeRamos SW. High rates of enteric protozoal infections in selected homosexual men attending a venereal disease clinic. *Sex Transm Dis.* 1978;5(4):155-157.

210. Markell EK, Kuritsubo RA. Intestinal protozoan infections: prevalence in the San Francisco Bay Area. *West J Med.* 1981;135(3):188-190.

211. Allason-Jones E, Mindel A, Sargeaunt P, Williams P. *Entamoeba histolytica* as a commensal intestinal parasite in homosexual men. *N Engl J Med.* 1986;315(6):353-356.

212. Sargeaunt PG, Williams JE. Electrophoretic isoenzyme patterns of the pathogenic and non-pathogenic intestinal amoebae of man. *Trans R Soc Trop Med Hyg.* 1979;73(2):225-227.

213. Quinn TC, Stamm WE, Goodell SE, et al. The polymicrobial origin of intestinal infections in homosexual men. *N Engl J Med.* 1983;309(10):576-582.

214. Laughon BE, Druckman DA, Vernon A, et al. Prevalence of enteric pathogens in homosexual men with and without acquired immunodeficiency syndrome. *Gastroenterology.* 1988;94(4):984-993.

215. Pakianathan MR, McMillan A. Intestinal protozoa in homosexual men in Edinburgh. *Int J STD AIDS.* 1999;10(12):780-784.

216. Di Benedetto MA, Di Piazza F, Amodio E, Taormina S, Romano N, Firenze A. Prevalence of sexually transmitted infections and enteric protozoa among homosexual men in western Sicily (south Italy). *J Prev Med Hyg.* 2012;53(4):181-185.

217. Anderson R, May R. *Infectious Diseases of Humans: Dynamics and Control.* Oxford: Oxford University Press. 1992.

218. Ghani AC, Swinton J, Garnett GP. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex Transm Dis.* 1997;24(1):45-56.

219. Garnett GP. The geographical and temporal evolution of sexually transmitted disease epidemics. *Sex Transm Infect.* 2002;78 Suppl 1:i14-19.

220. Doherty IA, Padian NS, Marlow C, Aral SO. Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections. *J Infect Dis.* 2005;191 Suppl 1:S42-54.

221. Anderson RM, May RM. Epidemiological parameters of HIV transmission. *Nature.* 1988;333(6173):514-519.

222. Escobedo AA, Almirall P, Alfonso M, Cimerman S, Chacin-Bonilla L. Sexual transmission of giardiasis: A neglected route of spread? *Acta Tropica.* 2014;132:106-111.

223. Savage EJ, Mohammed H, Leong G, Duffell S, Hughes G. Improving surveillance of sexually transmitted infections using mandatory electronic clinical reporting: the genitourinary medicine clinic activity dataset, England, 2009 to 2013. *Euro Surveill.* 2014;19(48):20981.

224. Mercer CH, Aicken CR, Tanton C, et al. Serial monogamy and biologic concurrency: measurement of the gaps between sexual partners to inform targeted strategies. *Am J Epidemiol.* 2013;178(2):249-259.

225. Wayal S, Reid D, Blomquist PB, Weatherburn P, Mercer CH, Hughes G. The Acceptability and Feasibility of Implementing a Bio-Behavioral Enhanced Surveillance Tool for Sexually Transmitted Infections in England: Mixed-Methods Study. *JMIR Public Health Surveill.* 2018;4(2):e52.

226. Sewell J, Speakman A, Phillips AN, et al. A Cross-Sectional Study on Attitudes to and Understanding of Risk of Acquisition of HIV: Design, Methods and Participant Characteristics. *JMIR Res Protoc.* 2016;5(2):e58.

227. Speakman A, Rodger A, Phillips AN, et al. The 'Antiretrovirals, Sexual Transmission Risk and Attitudes' (ASTRA) study. Design, methods and participant characteristics. *PLoS One.* 2013;8(10):e77230.

228. Aicken CR, Cassell JA, Estcourt CS, et al. Rationale and development of a survey tool for describing and auditing the composition of, and flows between, specialist and community clinical services for sexually transmitted infections. *BMC Health Serv Res.* 2011;11:30.

229. Whitlock GG, Gibbons DC, Longford N, Harvey MJ, McOwan A, Adams EJ. Rapid testing and treatment for sexually transmitted infections improve patient care and yield public health benefits. *Int J STD AIDS.* 2018;29(5):474-482.

230. Information Commissioner's Office. *Anonymisation: managing data protection risk code of practice.* Cheshire: Information Commissioner's Office. 2012.

231. NHS Data Model and Dictionary for England. GUMCAD. Available from: https://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/genitourinary_medicine_clinic_activity_data_set_fr.asp?shownav=1. [Accessed 5th May 2018].

232. Public Health England. GUMCAD Sexual Health and HIV Activity Property Type (SHHAPT) code look-up and summary of definitions. Available from: https://www.gov.uk/government/publications/gumcad-clinical-guidelines [Accessed 5th May 2018].

233. Department for Communities and Local Government. The English Indices of Deprivation 2015: Statistical Release. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465791/English_Indices_of_Deprivation_2015_-_Statistical_Release.pdf [Accessed 30th July 2019].

234. Office for National Statistics. Mid-2017 Population Estimates for Lower Layer Super Output Areas in England and Wales by single year of age and sex - supporting information. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates [Accessed 30th July 2019].

235. Public Health England. PrEP Impact Trial: A pragmatic health technology assessment of PrEP and implementation. Protocol Number SSCR104, version 5.0 (25th July 2019). Available from: https://www.prepimpacttrial.org.uk/protocol [Accessed 9th October 2019].

236. Murphy NM, McLauchlin J, Ohai C, Grant KA. Construction and evaluation of a microbiological positive process internal control for PCR-based examination of food samples for *Listeria monocytogenes* and *Salmonella enterica*. *Int J Food Microbiol.* 2007;120(1-2):110-119.

237. Nair S, Ashton P, Doumith M, et al. WGS for surveillance of antimicrobial resistance: a pilot study to detect the prevalence and mechanism of resistance to azithromycin in a UK population of non-typhoidal *Salmonella*. *J Antimicrob Chemother.* 2016;71(12):3400-3408.

238. Nguyen MCP, Woerther PL, Bouvet M, Andremont A, Leclercq R, Canu A. *Escherichia coli* as reservoir for macrolide resistance genes. *Emerg Infect Dis.* 2009;15(10):1648-1650.

239. Hughes G, Nichols T, Peters L, Bell G, Leong G, Kinghorn G. Repeat infection with gonorrhoea in Sheffield, UK: predictable and preventable? *Sex Transm Infect.* 2013;89(1):38-44.

240. Desai S, Nardone A, Hughes G, et al. HIV incidence in an open national cohort of men who have sex with men attending sexually transmitted infection clinics in England. *HIV Med.* 2017;18(9):615-622.

241. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.

242. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol.* 2017;9:157-166.

243. Carpenter JR, Kenward MG. *Multiple imputation and its application.* Chichester: John Wiley & Sons Ltd. 2013.

244. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* 2006;59(10):1087-1091.

245. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol.* 1995;142(12):1255-1264.

246. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30:377-399.

247. Mohammed H, Nardone A, Gilbart V, Desai S, Hughes G. Monitoring STI risk behaviour and partner notification outcomes through routine national surveillance: a pilot study in England. In: 2014 STD Prevention Conference, June 9-12, Atlanta, USA. 2014.

248. Erens B, Phelps A, Clifton S, et al. Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *Sex Transm Infect.* 2014;90(2):84-89.

249. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol.* 2004;159(7):702-706.

250. Kotar T, Pirs M, Steyer A, et al. Evaluation of rectal swab use for the determination of enteric pathogens: a prospective study of diarrhoea in adults. *Clin Microbiol Infect.* 2019;25(6):733-738.

251. Fenton KA, Johnson AM, McManus S, Erens B. Measuring sexual behaviour: methodological challenges in survey research. *Sex Transm Infect.* 2001;77(2):84-92.

252. Kirkwood BR, Sterne JAC. *Essential Medical Statistics (2nd Edition).* Oxford: Blackwell Publishing. 2003.

253. Tam CC, Higgins CD, Neal KR, et al. Chicken consumption and use of acid-suppressing medications as risk factors for *Campylobacter* enteritis, England. *Emerg Infect Dis.* 2009;15(9):1402-1408.

254. Tam CC, Rodrigues LC, O'Brien SJ. The study of infectious intestinal disease in England: what risk factors for presentation to general practice tell us about potential

for selection bias in case-control studies of reported cases of diarrhoea. *Int J Epidemiol.* 2003;32(1):99-105.

255. Gillespie IA, O'Brien SJ, Adak GK, Ward LR, Smith HR. Foodborne general outbreaks of *Salmonella Enteritidis* phage type 4 infection, England and Wales, 1992-2002: where are the risks? *Epidemiol Infect.* 2005;133(5):795-801.

256. Williamson DA, Chow EPF, Lee D, et al. Risk Factors for Asymptomatic Enteric Pathogen Detection Among Men Who Have Sex With Men. *Open Forum Infect Dis.* 2019;6(9):ofz326.

257. Curtis TJ, Rodger AJ, Burns F, Nardone A, Copas A, Wayal S. Patterns of sexualised recreational drug use and its association with risk behaviours and sexual health outcomes in men who have sex with men in London, UK: a comparison of cross-sectional studies conducted in 2013 and 2016. *Sex Transm Infect.* 2020;96(3):197-203.

258. Wayal S, Reid D, Weatherburn P, et al. Association between knowledge, risk behaviours, and testing for sexually transmitted infections among men who have sex with men: findings from a large online survey in the United Kingdom. *HIV Med.* 2019;20(8):523-533.

259. Mitchell HD, Desai S, Mohammed H, et al. Preparing for PrEP: estimating the size of the population eligible for HIV pre-exposure prophylaxis among men who have sex with men in England. *Sex Transm Infect.* 2019;95(7):484-487.

260. Desai S, Burns F, Schembri G, et al. Sexual behaviours and sexually transmitted infection outcomes in a cohort of HIV-negative men who have sex with men attending sexual health clinics in England. *Int J STD AIDS.* 2018;29(14):1407-1416.

261. Wayal S, Hart G, Copas A, Edwards S, Cassell JA. Sexual behaviour, partnership patterns and STI diagnoses among HIV positive MSM: Implications for HIV/STIs transmission and partner notification. *Sex Transm Infect.* 2012;88(Suppl 1):A7.

262. Mohammed H, Furegato M, Hughes G. Inequalities in sexually transmitted infection risk among black and minority ethnic men who have sex with men in England. *Sex Transm Infect.* 2016;92(Suppl 1):A42.

263. Penders J, Stobberingh EE, Savelkoul PH, Wolffs PF. The human microbiome as a reservoir of antimicrobial resistance. *Front Microbiol.* 2013;4:87.

264. Salyers AA, Gupta A, Wang Y. Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends Microbiol.* 2004;12(9):412-416.

265. Nys S, Okeke IN, Kariuki S, Dinant GJ, Driessen C, Stobberingh EE. Antibiotic resistance of faecal *Escherichia coli* from healthy volunteers from eight developing countries. *J Antimicrob Chemother.* 2004;54(5):952-955.

266. Gomes C, Ruiz-Roldan L, Mateu J, Ochoa TJ, Ruiz J. Azithromycin resistance levels and mechanisms in *Escherichia coli. Sci Rep.* 2019;9(1):6089.

267. Public Health England. *Identification of Shigella species*. UK Standards for Microbiology Investigations (ID20 Issue 3). London: Public Health England. 2015. Contains public sector information licensed under the Open Government Licence v3.0.

268. Gentle A, Ashton PM, Dallman TJ, Jenkins C. Evaluation of Molecular Methods for Serotyping *Shigella flexneri*. *J Clin Microbiol.* 2016;54(6):1456-1461.

269. Chattaway MA, Greig DR, Gentle A, Hartman HB, Dallman TJ, Jenkins C. Whole-Genome Sequencing for National Surveillance of *Shigella flexneri*. *Front Microbiol.* 2017;8:1700.

270. Illumina. An introduction to Next-Generation Sequencing Technology. Available from: https://www.illumina.com [Accessed 22nd April 2018].

271. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* 2012;13(9):601-612.

272. Field N, Cohen T, Struelens MJ, et al. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect Dis.* 2014;14(4):341-352.

273. Dallman T, Ashton P, Schafer U, et al. SnapperDB: a database solution for routine sequencing analysis of bacterial isolates. *Bioinformatics.* 2018;34(17):3028-3029.

274. Chattaway MA, Schaefer U, Tewolde R, Dallman TJ, Jenkins C. Identification of *Escherichia coli* and *Shigella* Species from Whole-Genome Sequences. *J Clin Microbiol.* 2017;55(2):616-623.

275. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120.

276. Ashton P, Nair S, Peters T, et al. Revolutionising Public Health Reference Microbiology using Whole Genome Sequencing: *Salmonella* as an exemplar. . *bioRxiv.* 2015;033225.

277. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-359.

278. Doumith M, Godbole G, Ashton P, et al. Detection of the plasmid-mediated mcr-1 gene conferring colistin resistance in human and food isolates of *Salmonella enterica* and *Escherichia coli* in England and Wales. *J Antimicrob Chemother.* 2016;71(8):2300-2305.

279. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* 2013;1303(3997).

280. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.

281. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-1303.

282. Jenkins C, Dallman TJ, Launders N, et al. Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing *Escherichia coli* O157 Associated with Consumption of Watercress. *Appl Environ Microbiol.* 2015;81(12):3946-3952.

283. McDonnell J, Dallman T, Atkin S, et al. Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of *Shigella sonnei* in the UK. *Epidemiol Infect.* 2013;141(12):2568-2575.

284. Rew V, Mook P, Trienekens S, et al. Whole-genome sequencing revealed concurrent outbreaks of shigellosis in the English Orthodox Jewish Community caused by multiple importations of *Shigella sonnei* from Israel. *Microb Genom.* 2018;4(3).

285. Butcher H, Elson R, Chattaway MA, et al. Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk. *Epidemiol Infect.* 2016;144(13):2812-2823.

286. Inns T, Ashton PM, Herrera-Leon S, et al. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. *Epidemiol Infect.* 2017;145(2):289-298.

287. Dallman T, Inns T, Jombart T, et al. Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. *Microb Genom.* 2016;2(8):e000070.

288. Waldram A, Dolan G, Ashton PM, Jenkins C, Dallman TJ. Epidemiological analysis of *Salmonella* clusters identified by whole genome sequencing, England and Wales 2014. *Food Microbiol.* 2018;71:39-45.

289. Joloza T, Evans J, O'Brien R, Potter-Collins A. *Measuring Sexual Identity: An Evaluation Report.* Newport: Office for National Statistics. 2010.

290. Public Health England. HIV surveillance systems. Available from: https://www.gov.uk/guidance/hiv-surveillance-systems [Accessed 16th November 2019].

291. Mortimer JY, Salathiel JA. 'Soundex' codes of surnames provide confidentiality and accuracy in a national HIV database. *Commun Dis Rep CDR Rev.* 1995;5(12):R183-186.

292. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* 2015;43(3):e15.

293. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312-1313.

294. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(W1):W242-245.

295. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019.

296. Bouckaert R, Heled J, Kuhnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10(4):e1003537.

297. Connor TR, Barker CR, Baker KS, et al. Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri. Elife.* 2015;4:e07335.

298. Public Health England. *Producing modelled estimates of the size of the lesbian, gay and bisexual (LGB) population of England*. 2017. London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0.

299. Garrett V, Bornschlegel K, Lange D, et al. A recurring outbreak of *Shigella sonnei* among traditionally observant Jewish children in New York City: the risks of daycare and household transmission. *Epidemiol Infect.* 2006;134(6):1231-1236.

300. Baker KS, Dallman TJ, Behar A, et al. Travel- and Community-Based Transmission of Multidrug-Resistant *Shigella sonnei* Lineage among International Orthodox Jewish Communities. *Emerging Infectious Diseases.* 2016;22(9):1545-1553.

301. Reback CJ, Larkins S. Maintaining a heterosexual identity: sexual meanings among a sample of heterosexually identified men who have sex with men. *Arch Sex Behav.* 2010;39(3):766-773.

302. Geary RS, Tanton C, Erens B, et al. Sexual identity, attraction and behaviour in Britain: The implications of using different dimensions of sexual orientation to estimate the size of sexual minority populations and inform public health interventions. *PLoS One.* 2018;13(1):e0189607.

303. McCrickard LS, Crim SM, Kim S, Bowen A. Disparities in severe shigellosis among adults - Foodborne diseases active surveillance network, 2002-2014. *BMC Public Health.* 2018;18(1):221.

304. Mehandru S, Poles MA, Tenner-Racz K, et al. Lack of mucosal immune reconstitution during prolonged treatment of acute and early HIV-1 infection. *PLoS Med.* 2006;3(12):e484.

305. Evering TH, Mehandru S, Racz P, et al. Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. *PLoS Pathog.* 2012;8(2):e1002506.

306. Dallman TJ, Ashton PM, Byrne L, et al. Applying phylogenomics to understand the emergence of Shiga-toxin-producing *Escherichia coli* O157:H7 strains causing severe human disease in the UK. *Microb Genom.* 2015;1(3):e000029.

307. Glynn JR, Bradley DJ. The relationship between infecting dose and severity of disease in reported outbreaks of *Salmonella* infections. *Epidemiol Infect.* 1992;109(3):371-388.

308. Noguera-Julian M, Rocafort M, Guillen Y, et al. Gut Microbiota Linked to Sexual Preference and HIV Infection. *EBioMedicine.* 2016;5:135-146.

309. Kelley CF, Kraft CS, de Man TJ, et al. The rectal mucosa and condomless receptive anal intercourse in HIV-negative MSM: implications for HIV transmission and prevention. *Mucosal Immunol.* 2017;10(4):996-1007.

310. Armstrong AJS, Shaffer M, Nusbacher NM, et al. An exploration of Prevotella-rich microbiomes in HIV and men who have sex with men. *Microbiome.* 2018;6(1):198.

311. Palmer CD, Tomassilli J, Sirignano M, et al. Enhanced immune activation linked to endotoxemia in HIV-1 seronegative MSM. *AIDS.* 2014;28(14):2162-2166.

312. Ragonnet-Cronin M, Hue S, Hodcroft EB, et al. Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV.* 2018;5(6):e309-e316.

313. Public Health England. *Addressing the increase in syphilis in England: PHE Action Plan*. London: Public Health England. 2019. Contains public sector information licensed under the Open Government Licence v3.0.

314. Public Health England. *Laboratory surveillance of non-travel associated Shigella spp. infection in adult males, England: 2004 to 2017*. London: Public Health England. 2017. Contains public sector information licensed under the Open Government Licence v3.0.

315. Fenton KA, Imrie J. Increasing rates of sexually transmitted diseases in homosexual men in Western europe and the United States: why? *Infect Dis Clin North Am.* 2005;19(2):311-331.

316. Kanagarajah S, Waldram A, Dolan G, et al. Whole genome sequencing reveals an outbreak of *Salmonella* Enteritidis associated with reptile feeder mice in the United Kingdom, 2012-2015. *Food Microbiol.* 2018;71:32-38.

317. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology (Third edition)*. Philadelphia: Wolters Kluwer Health/Lippencott Williams and Wilkins. 2008.

318. Buchwald DS, Blaser MJ. A review of human salmonellosis: II. Duration of excretion following infection with nontyphi Salmonella. *Rev Infect Dis.* 1984;6(3):345-356.

319. Levine MM, Robins-Browne RM. Factors that explain excretion of enteric pathogens by persons without diarrhea. *Clin Infect Dis.* 2012;55 Suppl 4:S303-311.

320. Levine MM, DuPont HL, Khodabandelou M, Hornick RB. Long-term Shigella-carrier state. *N Engl J Med.* 1973;288(22):1169-1171.

321. Christopher PR, David KV, John SM, Sankarapandian V. Antibiotic therapy for Shigella dysentery. *Cochrane Database Syst Rev.* 2010(8):CD006784.

322. Havelaar AH, van Pelt W, Ang CW, et al. Immunity to *Campylobacter*: its role in risk assessment and epidemiology. *Crit Rev Microbiol.* 2009;35(1):1-22.

323. Watts C, Zimmerman C, Foss AM, Hossain M, Cox A, Vickerman P. Remodelling core group theory: the role of sustaining populations in HIV transmission. *Sex Transm Infect.* 2010;86 Suppl 3:iii85-92.

324. Pines HA, Gorbach PM, Weiss RE, et al. Sexual risk trajectories among MSM in the United States: implications for pre-exposure prophylaxis delivery. *J Acquir Immune Defic Syndr.* 2014;65(5):579-586.

325. Datta J, Reid D, Hughes G, Mercer CH, Wayal S, Weatherburn P. Awareness of and attitudes to sexually transmissible infections among gay men and other men who have sex with men in England: a qualitative study. *Sex Health.* 2019;16(1):18-24.

326. Metcalfe R, Laird G, Nandwani R. Don't ask, sometimes tell. A survey of men who have sex with men sexual orientation disclosure in general practice. *Int J STD AIDS.* 2015;26(14):1028-1034.

327. Bains M, Crook P, Field N, Hughes G. Safer chemsex: Consideration of *Shigella*. *British Journal of General Practice.* 2016;66(644):124.

328. NHS England. Implementation Guidance Fundamental Standard for Sexual Orientation Monitoring. Available from: https://www.england.nhs.uk/wp-content/uploads/2017/10/implementation-guidance-fundamental-standard-for-sexual-orientation-monitoring.pdf [Accessed 10th December 2019]. 2017.

329. NHS England. Sexual Orientation Monitoring Information Standard. Available from: https://www.england.nhs.uk/about/equality/equality-hub/sexual-orientation-monitoring-information-standard/ [Accessed 10th December 2019].

330. MacArthur RD, DuPont HL. Etiology and pharmacologic management of noninfectious diarrhea in HIV-infected individuals in the highly active antiretroviral therapy era. *Clin Infect Dis.* 2012;55(6):860-867.

331. Mercer CH, Prah P, Field N, et al. The health and well-being of men who have sex with men (MSM) in Britain: Evidence from the third National Survey of Sexual Attitudes and Lifestyles (Natsal-3). *BMC Public Health.* 2016;16:525.

332. Fifer H, Cole M, Hughes G, et al. Sustained transmission of high-level azithromycin-resistant *Neisseria gonorrhoeae* in England: an observational study. *Lancet Infect Dis.* 2018;18(5):573-581.

333. Fifer H, Saunders J, Soni S, Sadiq ST, FitzGerald M. 2018 UK national guideline for the management of infection with *Neisseria gonorrhoeae*. *Int J STD AIDS.* 2020;31(1):4-15.

334. Hughes G, Saunders J. *Mycoplasma genitalium*: the next sexually transmitted superbug? *BMJ.* 2018;363:k4376.

335. Lau A, Bradshaw CS, Lewis D, et al. The Efficacy of Azithromycin for the Treatment of Genital *Mycoplasma genitalium*: A Systematic Review and Meta-analysis. *Clin Infect Dis.* 2015;61(9):1389-1399.

336. Grant JS, Stafylis C, Celum C, et al. Doxycycline prophylaxis for bacterial sexually transmitted infections. *Clin Infect Dis.* 2019.

337. Molina JM, Charreau I, Chidiac C, et al. Post-exposure prophylaxis with doxycycline to prevent sexually transmitted infections in men who have sex with men: an open-label randomised substudy of the ANRS IPERGAY trial. *Lancet Infect Dis.* 2018;18(3):308-317.

338. Bolan RK, Beymer MR, Weiss RE, Flynn RP, Leibowitz AA, Klausner JD. Doxycycline prophylaxis to reduce incident syphilis among HIV-infected men who have sex with men who continue to engage in high-risk sex: a randomized, controlled pilot study. *Sex Transm Dis.* 2015;42(2):98-103.

339. McCall H, Foster F, Hughes G, Oxenham C, Field N. Let's talk about sex: sexual health history taking in health protection. In: Public Health England Annual Conference 2017, September 12-13, Warwick, UK. 2017.

340. Newman CE, Persson A, Holt M, et al. Understanding expert views on defining and reaching heterosexually-identified men who have sex with men for health promotion and care (short report). Available from: https://csrh.arts.unsw.edu.au/research/projects/straight-msm/ [Accessed 10th December 2019]. 2016.

341. Newman CE, Persson A, Manolas P, et al. "So Much Is at Stake": Professional Views on Engaging Heterosexually Identified Men who Have Sex with Men with Sexual Health Care in Australia. *Sex Res Soc Policy.* 2018;15(3):302-311.

342. Centers for Disease Control and Prevention. PulseNet. https://www.cdc.gov/pulsenet/index.html [Accessed 10th March 2020].

343. Carleton H, Gerner-Smidt P. Whole-genome sequencing is taking over foodborne disease surveillance. *Microbe.* 2016;11:311-317.

344. Nadon C, Van Walle I, Gerner-Smidt P, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill.* 2017;22(23).

Appendices

Appendix for Chapter 2

Appendix 2.1: Search terms for literature review

Relevant studies were identified through a scoping search of the literature using the electronic databases MEDLINE (OvidSP), EMBASE (OvidSP), Web of Science Core Collection and Scopus. Each database was searched from inception until 23rd July 2018. The search strategy comprised of Medical Subject Headings (MeSH) and text word searches. MeSH terms were exploded where appropriate to broaden the search. The initial search terms were piloted in MEDLINE prior to selection. Additional terms identified during the pilot search were incorporated into the final search strategy and those that did not offer further literature were dropped. The search terms consisted of specific GI infections and associated symptoms (e.g. diarrhoea), sexual transmission, infectious disease outbreaks and MSM. These were combined using Boolean operators ("AND" and "OR"). The GI infection terms were selected to represent the broad range of bacterial, viral and protozoal pathogens that can be transmitted through sexual contact. The reference lists of selected articles were scanned to identify any potential publications that may have been missed through the search strategy. Grey literature including surveillance reports were also included in the narrative review. The search strategy for Medline can be found below.

1	Gastroenteritis/
2	acute gastroenteritis.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
3	dysentery.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
4	diarrh?ea*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
5	exp diarrhea/

6	enteric infect*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
7	enteric pathogen*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
8	gastrointestinal infectio*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
9	gastrointestinal pathogen*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
10	Dysentery, Bacillary/
11	infectious intestinal disease*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
12	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13	exp Sexually Transmitted Diseases/
14	sexual* transmi*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
15	(sti or std or venereal).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
16	(outbreak* or epidemic*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
17	exp Disease Outbreaks/
18	13 or 14 or 15 or 16 or 17
19	exp *Shigella*/
20	(*Shigella* or shigellosis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

21	Campylobacter Infections/
22	exp Campylobacter/
23	(campylobacter or campylobacteri*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
24	Salmonella Infections/
25	Salmonella/
26	exp Salmonella enterica/
27	salmonella.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
28	salmonellosis.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
29	exp Escherichia coli/
30	Escherichia coli.mp. or E.coli. or E coli.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
31	(VTEC or STEC).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
32	exp Entamoeba/
33	entam?eba.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
34	giardia*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
35	Giardiasis/
36	Escherichia coli Infections/
37	Amebiasis/

38	Am?ebiasis.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
39	exp Giardia/
40	exp Hepatitis A virus/
41	Hepatitis A/
42	("Hepatitis A" or "Hep A").mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
43	Cryptosporidiosis/
44	exp Cryptosporidium/
45	exp Cryptosporidiiae/
46	cryptosporid*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
47	microsporidia/
48	microsporidiosis.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
49	(microsporidia or microsporidium).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
50	microsporidia, unclassified/
51	19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50
52	12 or 51
53	Homosexuality, Male/
54	Sexual Behavior/
55	(MSM or msm or "men who have sex with men").mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

56	gay.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
57	Homosexuality/
58	homosexual*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
59	bisexual*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
60	sex* behavio?r*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
61	sex* activit*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
62	53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61
63	18 and 52 and 62
64	limit 63 to english language

Appendix for Chapter 3

Appendix 3.1: Patient information poster

If you are having a rectal swab taken…

the left over sample may be included in a postgraduate student research study being carried out by researchers from University College London (UCL), Public Health England (PHE) and the Wellcome Trust Sanger Institute. The full study title is ***"Prevalence and characteristics of sexually transmissible enteric infections (STEIs) among men who have sex with men (MSM) attending a London sexual health clinic: a student study".***

This study has received funding from the National Institute for Health Research Health Protection Research Units (NIHR HPRUs) in Blood Borne and Sexually Transmitted Infections, and in Gastrointestinal Infections.

This study has been reviewed and approved by an independent NHS Research Ethics Committee and an NHS Research and Development department.

What is the study about?

There are a range of microbes (bugs) such as bacteria and viruses that live in the gut. Some bugs can make you sick; they can cause diarrhoea, stomach cramps and fever. They can be passed on by small traces of faeces (poo) getting into your mouth during and after sex. We are investigating the number and range of gut bugs present in gay, bisexual and other men who have sex with men (e.g. *Shigella*). This will help us to understand the extent to which these bugs are spreading and to identify who is at higher risk.

How common are these bugs in men who have sex with men?

There is currently little information available, but we think that between 1% and 9% of men who have sex with men may have a bug that can make you sick in their gut. We also think that some people might not have any symptoms. Testing and treatment for these bugs is not necessary unless you have symptoms as the bugs will clear up on their own. Our study will help to provide further information about these questions.

What will happen to my sample?

Once your clinical tests have been done, the left over sample will be sent to the laboratory at Public Health England. Testing for gut bugs will be done separately and only after anything that identifies you has been removed from the sample. The results will be linked to data routinely collected by Public Health England as part of national STI surveillance and to other information collected at the clinic.

Will the sample have my name on it?

No. The leftover sample will only have a patient number on it. No identifying information such as date of birth, postcode or name will be attached to the sample. Prior to testing, the patient number will be removed from the sample and replaced with an anonymous study number.

Will my sample be included?

We would like to include all rectal swabs from men aged 16 years and above attending 56 Dean Street or Dean Street Express. If you would prefer to have your sample excluded, please speak to a member of clinic staff and sign the opt-out log to make sure your sample is not used.

Will any extra tests be done?

No. Testing for gut bugs will only be done using the left over sample from your routine rectal swab.

Will I get the test results?

No – because the testing is being done anonymously. Testing for gut bugs will be done separately and the test results will only be associated with a study number. The results will not be sent back to the clinic. The results from these tests cannot be used for clinical diagnosis or decisions on the care or treatment you currently receive.

What should I do if I'm worried about gut symptoms?

If you are experiencing symptoms of diarrhoea, stomach cramps or fever, you should visit your GP. Alternatively, if you are visiting 56 Dean Street (the main clinic for people with symptoms), you could discuss your symptoms with the healthcare professional you see today. If you are visiting Dean Street Express today, you may be referred to 56 Dean Street.

What should I do if I have any questions or concerns about the study?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions on

If you remain unhappy and wish to complain formally, you can do this by contacting the Patient Advice and Liaison Service on 020 3315 6727 or email cwpals@chelwest.nhs.uk. Further details can be found here: http://www.chelwest.nhs.uk/your-visit/patient-advice

UCL Institute for Global Health Mortimer Market Centre, Capper Street, London WC1E 6JB
Sexually transmissible enteric infections in MSM Poster
V2.0 31/10/2017
IRAS project ID: 225176

Funded by NHS National Institute for Health Research

Appendix 3.2: Patient information leaflet

If you are having a rectal swab taken…

The leftover sample may be included in a postgraduate student research study being carried out by researchers from University College London (UCL), Public Health England (PHE) and the Wellcome Trust Sanger Institute.

This study has received funding from the National Institute for Health Research Health Protection Research Units (NIHR HPRUs) in Blood Borne and Sexually Transmitted Infections, and in Gastrointestinal Infections.

This study has been reviewed and approved by an independent NHS Research Ethics Committee and an NHS Research and Development department.

What should I do if I have any questions or concerns about the study?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions on

If you remain unhappy and wish to complain formally, you can do this by contacting the Patient Advice and Liaison Service on 020 3315 6727 or email cwpals@chelwest.nhs.uk.

Further details can be found here: http://www.chelwest.nhs.uk/your-visit/patient-advice

UCL Institute for Global Health
Mortimer Market Centre
Off Caper Street
London, WC1E 6JB

Sexually transmissible enteric infections in MSM Leaflet
V1.0 31/10/2017
IRAS project ID: 225176

"Prevalence and characteristics of sexually transmissible enteric infections (STEIs) among men who have sex with men (MSM) attending a London sexual health clinic: a student study"

Research study: Information for men who are having a rectal swab taken today

What is the study about?

There are a range of microbes (bugs) such as bacteria and viruses that live in the gut. Some bugs can make you sick; they can cause diarrhoea, stomach cramps and fever. They can be passed on by small traces of faeces (poo) getting into your mouth during and after sex. We are investigating the number and range of gut bugs present in gay, bisexual and other men who have sex with men (e.g. *Shigella*). This will help us to understand the extent to which these bugs are spreading and to identify who is at higher risk.

How common are these bugs in men who have sex with men?

There is currently little information available, but we think that between 1% and 9% of men who have sex with men may have a bug that can make you sick in their gut. We also think that some people might not have any symptoms. Testing and treatment for these bugs is not necessary unless you have symptoms as the bugs will clear up on their own. Our study will help to provide further information about these questions.

What will happen to my sample?

Once your clinical tests have been done, the left over sample will be sent to the laboratory at Public Health England. Testing for gut bugs will be done separately and only after anything that identifies you has been removed from the sample. The results will be linked to data routinely collected by Public Health England as part of national STI surveillance and to other information collected at the clinic.

Will the sample have my name on it?

No. The leftover sample will only have a patient number on it. No identifying information such as date of birth, postcode or name will be attached to the sample. Prior to testing, the patient number will be removed from the sample and replaced with an anonymous study number.

Will my sample be included?

We would like to include all rectal swabs from men aged 16 years and above attending 56 Dean Street or Dean Street Express. If you would prefer to have your sample excluded, please speak to a member of clinic staff and sign the opt-out log to make sure your sample is not used.

Will any extra tests be done?

No. Testing for gut bugs will only be done using the left over sample from your routine rectal swab.

Will I get the test results?

No – because the testing is being done anonymously. Testing for gut bugs will be done separately and the test results will only be associated with a study number. The results will not be sent back to the clinic. The results from these tests cannot be used for clinical diagnosis or decisions on the care or treatment you currently receive.

What should I do if I'm worried about gut symptoms?

If you are experiencing symptoms of diarrhoea, stomach cramps or fever, you should visit your GP. Alternatively, if you are visiting 56 Dean Street (the main clinic for people with symptoms), you could discuss your symptoms with the healthcare professional you see today. If you are visiting Dean Street Express today, you may be referred to 56 Dean Street.

Appendix 3.3: Opt-out study log

	Date specimen taken (dd/mm/yyyy)	**Clinic patient number**	**Patient signature**	**Staff initials**	**Details sent to PHE* (Y/N)**
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

*Date specimen taken and clinic patient number only

Appendix 3.4: Data management, linkage and anonymisation algorithm

The text below describes the data management, linkage and anonymisation algorithm used to process rectal specimens and data collected as part of the cross-sectional study at Dean Street. The text should be read whilst referring to Figure 3.1.

1. All residual rectal swabs were labelled with a barcoded clinic patient number and the date of attendance by clinic staff at Dean Street Express (DSE) or laboratory staff at North West London Pathology (NWLP). No other identifiable data (e.g. name, date of birth or postcode) were present. This is defined as pseud-anonymisation by the Information Commissioner's Office.

2. The study research nurse at Dean Street generated a list of clinic patient numbers and the date of attendance for individuals who opted out (C1) and I used this to discard the relevant samples. The C1 file was deleted after discarding the samples. For the remaining samples, I removed the clinic patient number from the specimen tube and replaced it with a unique study ID to anonymise the samples. An encrypted and password protected temporary electronic file (T1) was generated that contained the study ID, the clinic patient number and the date of attendance. The T1 file was sent to a surveillance scientist within the HIV/STI Department of PHE Colindale who retained it for the duration of the study. It was necessary to retain the T1 file to link the PCR results to socio-demographic, behavioural and clinical data from the clinic database and GUMCAD using the clinic patient number.

3. GUMCAD data are submitted to PHE six weeks after the end of each calendar quarter. This meant that there was a delay between specimen collection and data linkage. The T1 file was retained until the end of the study to enable linkage to future GUMCAD submissions to address any questions about STI co-infection (e.g. was the patient diagnosed with a bacterial STI at the study attendance date). I did not have access to T1 from this point onwards and it was stored securely on an encrypted drive at PHE, separately from other study files and the specimens. The surveillance scientist did not have access to the biological test results. All rectal swabs were tested after they were

anonymised with the unique study ID and after the T1 file was sent to the surveillance scientist. This provided an additional level of security to prevent intentional or unintentional identification of patients.

4. The research nurse at Dean Street sent the surveillance scientist an electronic file of clinic patient numbers for men aged <16 years old and for all women who had a rectal swab collected during the study period (C2), to enable ineligible samples to be identified and discarded. The surveillance scientist merged the C2 file with the T1 file to assign the study ID, and the clinic patient number was deleted from the merged file. The anonymised file containing the study ID was sent to me and the original C2 file and anonymised copy held by the surveillance scientist were deleted. The T1 file was updated to exclude the non-eligible samples.

5. I discarded all rectal swabs belonging to <16 year olds and women. This process ensured that rectal swabs from adult men only were included in the study and underwent DNA extraction. I created a biological dataset containing the study ID and PCR results only. I was not able to access the clinic patient number or link the PCR results to this number (see point 2 above).

6. Dean Street clinic staff sent the surveillance scientist clinical and behavioural data extracted from the Dean Street clinic database along with the clinic patient number (C3), for all rectal swabs collected from adult men during the study period. The surveillance scientist linked C3 file to the anonymous study ID, using T1 file, and removed the clinic patient number from the merged file. The anonymised clinic data was then sent to me. The original C3 file and the anonymised copy held by the surveillance scientist were deleted.

7. The surveillance scientist merged the GUMCAD data (2008 – 2018) with T1 using the clinic patient number and date of attendance. The use of historical GUMCAD data was included to enable identification of individuals with a prior STI diagnosis. Variables

were grouped into categories where appropriate to minimise the risk of deductive identification (e.g. age group rather than age). Once the clinic patient number was removed, the anonymised data containing the study ID was sent to me. The surveillance scientist deleted the merged GUMCAD data with the study ID. T1 was retained initially to facilitate further linkage to future GUMCAD data and deleted at the end of the study.

8. I merged all datasets using the study ID (biological results, clinic data and GUMCAD) to create an anonymous study dataset containing the biological PCR results, demographic, clinical and behavioural data. I did not have access to the clinic patient number to link the PCR results to an individual, and even then, the pseudo-anonymised nature of the clinic patient number did not allow identification of the patient by the research study team.

Appendix 3.5: Study approval documentation

Health Research Authority

Dr Nigel Field
Centre for Molecular Epidemiology and Translational Research, Institute for Global Health

Email: hra.approval@nhs.net

22 November 2017

Dear Dr Field

Letter of HRA Approval

Study title:	**Prevalence and characteristics of sexually transmissible enteric infections (STEIs) among men who have sex with men (MSM) attending a London sexual health clinic: a student study**
IRAS project ID:	**225176**
REC reference:	**17/LO/1722**
Sponsor	**University College London**

I am pleased to confirm that **HRA Approval** has been given for the above referenced study, on the basis described in the application form, protocol, supporting documentation and any clarifications noted in this letter.

Participation of NHS Organisations in England

The sponsor should now provide a copy of this letter to all participating NHS organisations in England.

Appendix B provides important information for sponsors and participating NHS organisations in England for arranging and confirming capacity and capability. **Please read *Appendix B* carefully**, in particular the following sections:

- *Participating NHS organisations in England* – this clarifies the types of participating organisations in the study and whether or not all organisations will be undertaking the same activities
- *Confirmation of capacity and capability* - this confirms whether or not each type of participating NHS organisation in England is expected to give formal confirmation of capacity and capability. Where formal confirmation is not expected, the section also provides details on the time limit given to participating organisations to opt out of the study, or request additional time, before their participation is assumed.
- *Allocation of responsibilities and rights are agreed and documented (4.1 of HRA assessment criteria)* - this provides detail on the form of agreement to be used in the study to confirm capacity and capability, where applicable.

Further information on funding, HR processes, and compliance with HRA criteria and standards is also provided.

IRAS project ID	225176

It is critical that you involve both the research management function (e.g. R&D office) supporting each organisation and the local research team (where there is one) in setting up your study. Contact details and further information about working with the research management function for each organisation can be accessed from the HRA website.

Appendices

The HRA Approval letter contains the following appendices:

- A – List of documents reviewed during HRA assessment
- B – Summary of HRA assessment

After HRA Approval

The document *"After Ethical Review – guidance for sponsors and investigators"*, issued with your REC favourable opinion, gives detailed guidance on reporting expectations for studies, including:

- Registration of research
- Notifying amendments
- Notifying the end of the study

The HRA website also provides guidance on these topics, and is updated in the light of changes in reporting expectations or procedures.

In addition to the guidance in the above, please note the following:

- HRA Approval applies for the duration of your REC favourable opinion, unless otherwise notified in writing by the HRA.
- Substantial amendments should be submitted directly to the Research Ethics Committee, as detailed in the *After Ethical Review* document. Non-substantial amendments should be submitted for review by the HRA using the form provided on the HRA website, and emailed to hra.amendments@nhs.net.
- The HRA will categorise amendments (substantial and non-substantial) and issue confirmation of continued HRA Approval. Further details can be found on the HRA website.

Scope

HRA Approval provides an approval for research involving patients or staff in NHS organisations in England.

If your study involves NHS organisations in other countries in the UK, please contact the relevant national coordinating functions for support and advice. Further information can be found through IRAS.

If there are participating non-NHS organisations, local agreement should be obtained in accordance with the procedures of the local participating non-NHS organisation.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application

procedure. If you wish to make your views known please use the feedback form available on the HRA website.

HRA Training

We are pleased to welcome researchers and research management staff at our training days – see details on the HRA website.

Your IRAS project ID is **225176**. Please quote this on all correspondence.

Yours sincerely

Aliki Sifostratoudaki
Assessor

Email: hra.approval@nhs.net

Copy to: *Ms Jessica Broni-Tabi, University College London, Sponsor Contact*
Mr Damon Foster, Chelsea and Westminster Hospital NHS Foundation Trust, R&D Contact

Appendix 3.6: PCR testing for bacterial gastrointestinal infections in rectal swabs taken from MSM: Primers, probes and gene targets for PCR assays

Causative agent	Target	Name	Sequence (5’ – 3’)	5’ mod	3’ mod	Gene targeted
Campylobacter jejuni	mapA	mapA Forward	CTGGTGGTTTTGAAGCAAAGATT			Gene specific for *C. Jejuni* (mapA)
	mapA	mapA Reverse	CAATACCAGTGTCTAAAGTGCGTTTAT			
	mapA	mapA Probe	TTGAATTCCAACATCGCTAATGTATAAAAGCCCTTT	FAM	BHQ1	
Campylobacter coli	ceuE	ceuC Forward	AAGCTCTTATTGTTCTAACCAATTCTAACA			Gene specific for *C. Coli* (ceuE)
	ceuE	ceuC Reverse	TCATCCACAGCATTGATTCCTAA			
	ceuE	ceuC Probe	TTGGACCTCAATCTCGCTTTGGAATCATT	Vic/YY	BHQ2	
Salmonella spp.	Ttr6	ttr Forward	CTCACCAGGAGATTACAACATGG			ttr
	Ttr4	ttr Reverse	AGCTCAGACCAAAAGTGACCATC			
	Ttr5	ttr Probe	CACCGACGGCGAGACCGACTTT	FAM	BHQ1	
STEC	stx1	stx1 Forward	GGATAATTTGTTTGCAGTTGATGTC			stx1
	stx1	stx1 Reverse	CAAATCCTGTCACATATAAATTATTTCGT			
	stx1	stx1 Probe	CGTAGATTATTAAACCGCCCTCCTCTGGA	Cy5	BHQ2	
	stx2	stx2 Forward	TTTGTYACTGTSACAGCWGAAGCYTTACG			stx2
	stx2	stx2 Reverse	CCCCAGTTCARWGTRAGRTCMACRTC			
	stx2	stx2 Probe	TCGTCAGGCACTGTCTGAAACTGCTCC	YY	BHQ2	

Causative agent	Target	Name	Sequence (5' – 3')	5' mod	3' mod	Gene targeted
EPEC	eae	eae Forward	CATTGATCAGGATTTTTCTGGTGATA			eae
	eae	eae Reverse	CTCATGCGGAAATAGCCGTTA			
	eae	eae Probe	ATAGTCTCGCCAGTATTCGCCACCAATACC	JOE	BHQ1	
EAEC	aagR	aagR Forward	CCATTTATCGCAATCAGATTAA			aggR
	aagR	aagR Reverse	CAAGCATCTACTTTTGATATTCC			
	aagR	aagR Probe	CAGCGATACATTAAGACGCCTAAAGGA	Cy5	BHQ2	
Shigella spp.	ipaH	ipaH Forward	AGGTCGCTGCATGGCTGGAA			ipaH
	ipaH	ipaH Reverse	CACGGTCCTCACAGCTCTCA			
	ipaH	ipaH Probe	AACTCAGTGCCTCTGCGGAGCTTCGACA	FAM	BHQ1	
Gfp *E. coli*	gfp	gfp Forward	CCTGTCCTTTTACCAGACAACCA			gfp
	gfp	gfp Reverse	GGTCTCTCTTTTCGTTGGGATCT			
	gfp	gfp Probe	TACCTGTCCACACAATCTGCCCTTTCG	Cy5	BHQ2	

STEC: Shiga toxin-producing *E. coli*, EAEC: Enteroaggregative *E. coli*, EPEC: Enteropathogenic *E. coli*

Appendix 3.7: Multiple imputation models for each variable with missing values

Variable to be imputed	Type of model	Auxiliary variables included in the model*
Age group	Ordinal logistic regression	Number of new sexual partners, bacterial STI in past year, HIV risk group, interest in specific high-risk practices, last condomless sex
Ethnic group	Multinomial logistic regression	Number of new sexual partners, region of birth, HIV risk group, interest in specific high-risk practices
Region of birth	Multinomial logistic regression	Number of new sexual partners, HIV risk group, last condomless sex, interest in specific high-risk practices
IMD quintile	Ordinal logistic regression	Region of birth
Sexual orientation	Logistic regression	HIV risk group, bacterial STI in past year
HIV risk group	Multinomial logistic regression	Number of new sexual partners, bacterial STI in past year, last condomless sex, ethnic group, region of birth, interest in specific high-risk practices
Interest in specific high-risk practices	Logistic regression	Number of new sexual partners, bacterial STI in past year, last condomless sex, ethnic group, HIV risk group
Number of sexual partners in past 3 months	Ordinal logistic regression	HIV risk group, bacterial STI in past year, ethnic group, region of birth, interest in specific high-risk practices, last condomless sex
Number of new sexual partners in past 3 months	Ordinal logistic regression	HIV risk group, bacterial STI in past year, ethnic group, region of birth, interest in specific high-risk practices, last condomless sex
Receptive anal sex in last 3 months	Logistic regression	Number of new sexual partners, HIV risk group, last condomless sex
Receptive oral sex in last 3 months	Logistic regression	Number of new sexual partners, HIV risk group, last condomless sex
Last condomless sex	Ordinal logistic regression	Number of new sexual partners, bacterial STI in past year, region of birth, interest in specific high-risk practices, HIV risk group
Gastrointestinal symptoms	Logistic regression	HIV risk group

*All models included HIV status, age, clinic and any BEP. Model for HIV risk group did not include HIV status. Imputation of HIV risk group in all models was conditional on HIV status. Missing values imputed using multiple imputation by chained equations (MICE) using the Stata command `mi impute chained`

Appendix 3.8: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen: sensitivity analysis using the missing indicator method

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic						
DSE	174/1709	10.2	1.00	0.210	1.00	0.138
56DS	33/407	8.1	0.80 (0.56-1.14)		0.76 (0.53-1.09)	
Age group						
16-24	19/272	7.0	1.00	0.173	1.00	0.268
25-34	98/1020	9.6	1.39 (0.86-2.23)	0.068[a]	1.35 (0.84-2.17)	0.112[a]
35+	90/824	10.9	1.56 (0.97-2.52)		1.49 (0.92-2.43)	
Ethnic group						
White	169/1576	10.7	1.00	0.129	1.00	0.198
Black	8/76	10.5	0.98 (0.50-1.92)		0.97 (0.50-1.89)	
Mixed	5/131	3.8	0.36 (0.15-0.85)		0.37 (0.15-0.88)	
Asian	10/112	8.9	0.83 (0.45-1.53)		0.86 (0.47-1.59)	
Other	10/130	7.7	0.72 (0.39-1.32)		0.72 (0.39-1.33)	
Missing	5/91	5.5	0.52 (0.22-1.22)		0.60 (0.25-1.42)	
Region of birth						
UK	90/953	9.4	1.00	0.138	1.00	0.231
Europe	65/581	11.2	1.18 (0.88-1.60)		1.18 (0.87-1.60)	
Rest of world	49/485	10.1	1.07 (0.77-1.49)		1.06 (0.76-1.47)	
Missing	3/97	3.1	0.33 (0.11-1.02)		0.38 (0.12-1.20)	
IMD quintile						
1-2 (Most deprived)	144/1407	10.2	1.00	0.804	1.00	0.671
3	33/374	8.8	0.86 (0.60-1.24)		0.86 (0.60-1.24)	
4-5 (Least deprived)	27/304	8.9	0.87 (0.59-1.28)		0.86 (0.58-1.28)	
Missing	3/31	9.7	0.95 (0.32-2.80)		1.42 (0.48-4.14)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Sexual orientation						
Gay	200/2057	10.0	1.00	0.439	1.00	0.627
Bisexual/heterosexual	5/25	6.3	0.63 (0.27-1.50)		0.66 (0.28-1.57)	
Missing	2/34	5.9	0.59 (0.15-2.28)		0.85 (0.23-3.18)	
HIV status						
Negative/unknown	163/1744	9.4	1.00	0.141	1.00	0.162
Living with HIV	44/372	11.8	1.27 (0.92-1.73)		1.26 (0.91-1.76)	
HIV risk group						
Negative/not known, not on PrEP	60/930	6.5	1.00	<0.001	1.00	<0.001
Negative/not known, on PrEP	74/547	13.5	2.10 (1.52-2.90)		2.05 (1.48-2.85)	
Positive	44/372	11.8	1.83 (1.27-2.65)		1.86 (1.26-2.75)	
Missing	29/267	10.9	1.68 (1.10-2.57)		1.76 (1.15-2.68)	
Bacterial STI diagnosed at attendance						
No/unknown	145/1632	8.9	1.00	0.010	1.00	0.010
Yes	62/484	12.8	1.44 (1.09-1.91)		1.45 (1.09-1.91)	
Bacterial STI diagnosed in last year						
No/unknown	103/1251	8.2	1.00	0.004	1.00	0.009
Yes	104/865	12.0	1.46 (1.13-1.89)		1.42 (1.09-1.85)	
Interest in specific high-risk practices						
No	97/1074	9.0	1.00	0.026	1.00	0.057
Yes	85/701	12.1	1.34 (1.02-1.77)		1.30 (0.98-1.73)	
Missing	25/341	7.3	0.81 (0.53-1.24)		0.81 (0.52-1.28)	
Number of sexual partners in last 3 months						
0-1	14/182	7.7	1.00	0.008	1.00	0.018
2-4	52/678	7.7	1.00 (0.57-1.76)		0.96 (0.55-1.71)	
5-9	47/440	10.7	1.39 (0.78-2.46)		1.33 (0.75-2.37)	
10+	57/402	14.2	1.84 (1.06-3.22)		1.75 (1.00-3.05)	
Missing	37/414	8.9	1.16 (0.64-2.10)		1.17 (0.64-2.13)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Number of new sexual partners in last 3 months						
0-1	28/442	6.3	1.00	<0.001	1.00	0.001
2-4	54/588	9.2	1.45 (0.93-2.25)		1.45 (0.94-2.24)	
5-9	45/347	13.0	2.05 (1.30-3.21)		2.02 (1.29-3.16)	
10+	38/245	15.5	2.45 (1.54-3.89)		2.40 (1.51-3.80)	
Missing	42/494	8.5	1.34 (0.85-2.13)		1.38 (0.86-2.24)	
Receptive anal sex in last 3 months						
No	7/92	7.6	1.00	0.772	1.00	0.876
Yes	180/1816	9.9	1.30 (0.63-2.69)		1.20 (0.58-2.47)	
Missing	20/208	9.6	1.26 (0.55-2.88)		1.23 (0.52-2.90)	
Receptive oral sex in last 3 months						
No	6/47	12.8	1.00	0.775	1.00	0.689
Yes	178/1827	9.7	0.76 (0.36-1.63)		0.73 (0.35-1.55)	
Missing	23/242	9.5	0.74 (0.32-1.73)		0.77 (0.32-1.88)	
Last condomless sex						
Never/more than 6 weeks	54/621	8.7	1.00	0.041	1.00	0.060
Within 6 weeks	99/986	10.0	1.15 (0.84-1.58)		1.13 (0.82-1.55)	
Within 72 hours	38/275	13.8	1.59 (1.08-2.35)		1.54 (1.04-2.28)	
Missing	16/234	6.8	0.79 (0.46-1.35)		0.77 (0.43-1.36)	
Gastrointestinal symptoms						
No	148/1553	9.5	1.00	0.621	1.00	0.405
Yes	5/36	13.9	1.46 (0.64-3.34)		1.79 (0.76-4.20)	
Missing	54/527	10.3	1.08 (0.80-1.45)		0.98 (0.72-1.33)	

N=2116. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. For age, missing values are replaced with a constant value, here the median age, and an additional indicator variable is added to the model to indicate which values are missing. Global p-values by Wald test or test for trend (a). aPRs and p-value presented for age group for ease of interpretation. Each factor has been adjusted in a separate model for *a priori* factors (age, clinic and HIV status).

Appendix 3.9: Associations of behavioural factors with the detection of any bacterial enteric pathogen: sensitivity analyses incorporating worst- and best-case scenarios (low values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
HIV risk group						
Negative/not known, not on PrEP	89/1197	7.4	1.00	<0.001	1.00	<0.001
Negative/not known, on PrEP	74/547	13.5	1.82 (1.36-2.43)		1.77 (1.31-2.38)	
Positive	44/372	11.8	1.59 (1.13-2.24)		1.60 (1.11-2.29)	
Interest in specific high-risk practices						
No	122/1415	8.6	1.00	0.011	1.00	0.025
Yes	85/701	12.1	1.41 (1.08-1.83)		1.36 (1.04-1.78)	
Number of sexual partners in last 3 months						
0-1	51/596	8.6	1.00	0.004	1.00	0.009
2-4	52/678	7.7	0.90 (0.62-1.30)	0.002[a]	0.87 (0.60-1.26)	0.012[a]
5-9	47/440	10.7	1.25 (0.86-1.82)		1.20 (0.83-1.75)	
10+	57/402	14.2	1.66 (1.16-2.37)		1.57 (1.09-2.26)	
Number of new sexual partners in last 3 months						
0-1	70/936	7.5	1.00	<0.001	1.00	<0.001
2-4	54/588	9.2	1.23 (0.87-1.72)	<0.001[a]	1.22 (0.87-1.71)	<0.001[a]
5-9	45/347	13.0	1.73 (1.22-2.47)		1.70 (1.19-2.42)	
10+	38/245	15.5	2.07 (1.43-3.00)		2.02 (1.38-2.94)	
Receptive anal sex in last 3 months						
No	27/300	9.0	1.00	0.624	1.00	0.859
Yes	180/1816	9.9	1.10 (0.75-1.62)		1.04 (0.68-1.58)	
Receptive oral sex in last 3 months						
No	29/289	10.0	1.00	0.877	1.00	0.572
Yes	178/1827	9.7	0.97 (0.67-1.41)		0.89 (0.59-1.34)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Last condomless sex						
Never/ more than 6 weeks ago	70/855	8.2	1.00	0.022	1.00	0.033
Within 6 weeks	99/986	10.0	1.23 (0.92-1.64)		1.21 (0.90-1.62)	
Within 72 hours	38/275	13.8	1.69 (1.16-2.45)		1.64 (1.13-2.39)	
Gastrointestinal symptoms						
No	202/2080	9.7	1.00	0.395	1.00	0.181
Yes	5/36	13.9	1.43 (0.63-3.26)		1.79 (0.76-4.21)	

N=2116. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend (a). Each factor adjusted in separate model for *a priori* factors (age, clinic and HIV status – not shown).

Appendix 3.10: Associations of behavioural factors with the detection of any bacterial enteric pathogen: sensitivity analyses incorporating worst- and best-case scenarios (high values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
HIV risk group						
Negative/not known, not on PrEP	60/930	6.5	1.00	<0.001	1.00	<0.001
Negative/not known, on PrEP	103/814	12.7	1.96 (1.45-2.66)		1.96 (1.44-2.66)	
Positive	44/372	11.8	1.83 (1.27-2.65)		1.86 (1.26-2.74)	
Interest in specific high-risk practices						
No	97/1074	9.0	1.00	0.239	1.00	0.305
Yes	110/1042	10.6	1.17 (0.90-1.52)		1.15 (0.88-1.51)	
Number of sexual partners in last 3 months						
0-1	14/182	7.7	1.00	0.061	1.00	0.064
2-4	52/678	7.7	1.00 (0.57-1.76)	0.010[a]	0.96 (0.55-1.70)	0.034[a]
5-9	47/440	10.7	1.39 (0.78-2.46)		1.33 (0.75-2.35)	
10+	94/816	11.5	1.50 (0.87-2.56)		1.47 (0.85-2.52)	
Number of new sexual partners in last 3 months						
0-1	28/442	6.3	1.00	0.014	1.00	0.014
2-4	54/588	9.2	1.45 (0.93-2.25)	0.009[a]	1.44 (0.93-2.23)	0.008[a]
5-9	45/347	13.0	2.05 (1.22-3.21)		2.00 (1.28-3.15)	
10+	80/739	10.8	1.71 (1.13-2.59)		1.76 (1.16-2.68)	
Receptive anal sex in last 3 months						
No	7/92	7.6	1.00	0.479	1.00	0.610
Yes	200/2024	9.9	1.30 (0.63-2.68)		1.21 (0.59-2.50)	
Receptive oral sex in last 3 months						
No	6/47	12.8	1.00	0.480	1.00	0.419
Yes	201/2069	9.7	0.76 (0.36-1.63)		0.73 (0.34-1.56)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Last condomless sex						
Never/more than 6 weeks ago	54/621	8.7	1.00	0.525	1.00	0.562
Within 6 weeks	99/986	10.0	1.15 (0.84-1.58)		1.15 (0.83-1.57)	
Within 72 hours	54/509	10.6	1.22 (0.85-1.75)		1.22 (0.84-1.77)	
Gastrointestinal symptoms						
No	148/1553	9.5	1.00	0.515	1.00	0.848
Yes	59/563	10.5	1.10 (0.83-1.46)		1.03 (0.77-1.38)	

N=2116. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend (a). Each factor adjusted in separate model for *a priori* factors (age, clinic and HIV status – not shown).

Appendix 3.11: Associations of partner number with the detection of any bacterial enteric pathogen: sensitivity analyses using <u>single value imputation</u>

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted Model PR (95% CI)	p-value
Number of sexual partners in last 3 months						
0-1	14/182	7.7	1.00	0.004	1.00	0.010
2-4	89/1092	8.2	1.06 (0.62-1.82)	<0.001[a]	1.04 (0.60-1.80)	0.006[a]
5-9	47/440	10.7	1.39 (0.78-2.46)		1.34 (0.75-2.37)	
10+	57/402	14.2	1.84 (1.06-3.22)		1.75 (1.00-3.06)	
Number of sexual partners in last 3 months						
0-4	103/1274	8.1	1.00	0.001	1.00	0.003
5+	104/842	12.4	1.53 (1.18-1.98)		1.48 (1.14-1.92)	
Number of sexual partners in last 3 months						
Median (IQR)	4 (3-6)		1.02 (1.01-1.03)	<0.001	1.02 (1.01-1.03)	<0.001
Number of <u>new</u> sexual partners in last 3 months						
0-1	28/442	6.3	1.00	<0.001	1.00	<0.001
2-4	96/1082	8.9	1.40 (0.93-2.10)	<0.001[a]	1.42 (0.94-2.14)	<0.001[a]
5-9	45/347	13.0	2.05 (1.30-3.21)		2.02 (1.29-3.16)	
10+	38/245	15.5	2.45 (1.54-3.89)		2.39 (1.51-3.79)	
Number of <u>new</u> sexual partners in last 3 months						
0-4	124/1524	8.1	1.00	<0.001	1.00	<0.001
5+	83/592	14.0	1.72 (1.33-2.24)		1.68 (1.29-2.19)	
Number of <u>new</u> sexual partners in last 3 months						
Median (IQR)	3 (2-5)		1.02 (1.01-1.03)	<0.001	1.02 (1.01-1.03)	<0.001

N=2116. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend ([a]). Missing values for partner number replaced with median value. Each factor adjusted in separate model for *a priori* factors (age, clinic and HIV status – not shown).

Appendix 3.12: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen: sensitivity analysis using multiple imputation by chained equations

	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic				
DSE	1.00	0.210	1.00	0.128
56DS	0.80 (0.56-1.14)		0.75 (0.52-1.09)	
Age group				
16-24	1.00	0.174	1.00	0.269
25-34	1.39 (0.86-2.22)		1.34 (0.84-2.16)	
35+	1.56 (0.97-2.52)		1.49 (0.92-2.42)	
Ethnic group				
White	1.00	0.123	1.00	0.149
Black	0.96 (0.49-1.88)		0.95 (0.48-1.86)	
Mixed	0.34 (0.14-0.82)		0.36 (0.15-0.85)	
Asian	0.81 (0.44-1.49)		0.84 (0.46-1.55)	
Other	0.69 (0.37-1.28)		0.69 (0.38-1.28)	
Region of birth				
UK	1.00	0.533	1.00	0.551
Europe	1.19 (0.88-1.61)		1.18 (0.87-1.61)	
Rest of world	1.07 (0.77-1.49)		1.06 (0.76-1.47)	
IMD quintile				
1-2 (Most deprived)	1.00	0.628	1.00	0.616
3	0.87 (0.61-1.24)		0.87 (0.61-1.24)	
4-5 (Least deprived)	0.87 (0.59-1.29)		0.86 (0.58-1.28)	

	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Sexual orientation				
Gay	1.00	0.304	1.00	0.355
Bisexual/heterosexual	0.64 (0.27-1.50)		0.67 (0.28-1.58)	
HIV status				
HIV-negative/unknown	1.00	0.141	1.00	0.157
Living with HIV	1.27 (0.92-1.73)		1.27 (0.91-1.76)	
HIV risk group				
HIV-negative/unknown HIV status, not on PrEP	1.00	<0.001	1.00	<0.001
HIV-negative, on PrEP	2.12 (1.55-2.88)		2.08 (1.52-2.84)	
Living with HIV	1.78 (1.24-2.55)		1.81 (1.24-2.64)	
Bacterial STI diagnosed at attendance				
No/unknown	1.00	0.010	1.00	0.009
Yes	1.44 (1.09-1.91)		1.45 (1.10-1.92)	
Bacterial STI diagnosed in past year				
No/unknown	1.00	0.004	1.00	0.009
Yes	1.46 (1.13-1.89)		1.42 (1.09-1.85)	
Interest in specific high-risk practices				
No	1.00	0.075	1.00	0.124
Yes	1.29 (0.98-1.70)		1.25 (0.94-1.67)	
Number of sexual partners in last 3 months				
0-1	1.00	0.004	1.00	0.008
2-4	1.06 (0.60-1.87)		1.04 (0.59-1.84)	
5-9	1.45 (0.82-2.57)		1.40 (0.79-2.50)	
10+	1.92 (1.10-3.36)		1.84 (1.05-3.24)	

	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Number of <u>new</u> sexual partners in last 3 months				
0-1	1.00	0.001	1.00	0.001
2-4	1.43 (0.92-2.22)		1.44 (0.93-2.23)	
5-9	1.96 (1.24-3.10)		1.95 (1.24-3.07)	
10+	2.38 (1.49-3.79)		2.36 (1.48-3.77)	
Receptive anal sex in last 3 months				
No	1.00	0.448	1.00	0.567
Yes	1.33 (0.64-2.77)		1.24 (0.59-2.59)	
Receptive oral sex in last 3 months				
No	1.00	0.528	1.00	0.479
Yes	0.78 (0.36-1.68)		0.76 (0.35-1.63)	
Last condomless sex				
Never/more than 6 weeks ago	1.00	0.105	1.00	0.135
Within 6 weeks	1.14 (0.84-1.57)		1.14 (0.83-1.56)	
Within 72 hours	1.52 (1.03-2.26)		1.49 (1.00-2.22)	
Gastrointestinal symptoms				
No/unknown	1.00	0.412	1.00	0.208
Yes	1.41 (0.62-3.23)		1.74 (0.74-4.11)	

N=2116. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. aPRs and p-value presented for age group for ease of interpretation. Each factor has been adjusted in separate model for *a priori* factors (age, clinic and HIV status). Complete analysis of imputed datasets performed using the Stata commands `mi estimate and mi test`.

Appendix 3.13: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in HIV-negative MSM: sensitivity analysis using the missing indicator method

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic						
DSE	140/1461	9.6	1.00	0.445	1.00	0.442
56DS	23/283	8.1	0.85 (0.56-1.29)		0.85 (0.56-1.29)	
Age group						
16-24	17/262	6.5	1.00	0.128	1.00	0.130
25-34	80/876	9.1	1.42 (0.86-2.35)	0.043[a]	1.41 (0.85-2.35)	0.044[a]
35+	66/606	10.9	1.68 (1.00-2.81)		1.67 (1.00-2.80)	
Ethnic group						
White	136/1304	10.4	1.00	0.133	1.00	0.198
Black	5/56	8.9	0.86 (0.37-2.01)		0.86 (0.37-2.02)	
Mixed	3/108	2.8	0.27 (0.09-0.82)		0.28 (0.09-0.85)	
Asian	8/96	8.3	0.80 (0.40-1.58)		0.83 (0.42-1.64)	
Other	7/103	6.8	0.65 (0.31-1.35)		0.66 (0.32-1.37)	
Missing	4/77	5.2	0.50 (0.19-1.31)		0.58 (0.22-1.54)	
Region of birth						
UK	71/802	8.9	1.00	0.268	1.00	0.383
Europe	51/477	10.7	1.21 (0.86-1.70)		1.21 (0.86-1.71)	
Rest of world	38/385	9.9	1.11 (0.77-1.62)		1.11 (0.76-1.61)	
Missing	3/80	3.8	0.42 (0.14-1.31)		0.50 (0.16-1.56)	
IMD quintile						
1-2 (Most deprived)	108/1149	9.4	1.00	0.998	1.00	0.912
3	29/313	9.3	0.99 (0.67-1.46)		0.98 (0.66-1.45)	
4-5 (Least deprived)	23/252	9.1	0.97 (0.63-1.49)		0.95 (0.62-1.47)	
Missing	3/30	10.0	1.06 (0.36-3.16)		1.45 (0.49-4.28)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Sexual orientation						
Gay	157/1641	9.6	1.00	0.442	1.00	0.608
Bisexual/heterosexual	5/76	6.6	0.69 (0.29-1.63)		0.69 (0.29-1.63)	
Missing	1/27	3.7	0.39 (0.06-2.67)		0.60 (0.09-4.02)	
PrEP use						
No	60/930	6.5	1.00	<0.001	1.00	<0.001
Yes	74/547	13.5	2.10 (1.52-2.90)		2.04 (1.47-2.83)	
Missing	29/267	10.9	1.68 (1.10-2.57)		1.74 (1.14-2.66)	
Bacterial STI diagnosed at attendance						
No/unknown	116/1357	8.6	1.00	0.031	1.00	0.027
Yes	47/387	12.1	1.42 (1.03-1.96)		1.43 (1.04-1.97)	
Bacterial STI diagnosed in last year						
No/unknown	85/1078	7.9	1.00	0.008	1.00	0.009
Yes	78/666	11.7	1.49 (1.11-1.99)		1.47 (1.10-1.97)	
Interest in specific high-risk practices						
No	72/941	7.7	1.00	0.003	1.00	0.004
Yes	72/560	12.9	1.68 (1.23-2.29)		1.66 (1.21-2.72)	
Missing	19/243	7.8	1.02 (0.63-1.67)		1.03 (0.63-1.67)	
Number of sexual partners in last 3 months						
0-1	12/148	8.1	1.00	0.006	1.00	0.010
2-4	35/570	6.1	0.76 (0.40-1.42)		0.73 (0.39-1.36)	
5-9	41/370	11.1	1.37 (0.74-2.53)		1.30 (0.70-2.39)	
10+	44/328	13.4	1.65 (0.90-3.04)		1.55 (0.84-2.86)	
Missing	31/328	9.5	1.17 (0.62-2.21)		1.12 (0.59-2.12)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Number of new sexual partners in last 3 months						
0-1	18/355	5.1	1.00	<0.001	1.00	<0.001
2-4	41/505	8.1	1.60 (0.94-2.74)		1.58 (0.93-2.70)	
5-9	38/286	13.3	2.62 (1.53-4.49)		2.58 (1.51-4.40)	
10+	32/205	15.6	3.08 (1.77-5.34)		2.98 (1.72-5.16)	
Missing	34/393	8.7	1.71 (0.98-2.97)		1.68 (0.96-2.95)	
Receptive anal sex in last 3 months						
No	5/85	5.9	1.00	0.299	1.00	0.244
Yes	143/1537	9.3	1.58 (0.67-3.76)		1.53 (0.65-3.61)	
Missing	15/122	12.3	2.09 (0.79-5.53)		2.16 (0.82-5.69)	
Receptive oral sex in last 3 months						
No	4/42	9.5	1.00	0.571	1.00	0.366
Yes	141/1548	9.1	0.96 (0.37-2.46)		0.93 (0.36-2.39)	
Missing	18/154	11.7	1.23 (0.44-3.43)		1.31 (0.46-3.74)	
Last condomless sex						
Never/more than 6 weeks ago	49/566	8.7	1.00	0.051	1.00	0.062
Within 6 weeks	73/811	9.0	1.04 (0.74-1.47)		1.04 (0.74-1.47)	
Within 72 hours	32/228	14.0	1.62 (1.07-2.46)		1.61 (1.06-2.45)	
Missing	9/139	6.5	0.75 (0.38-1.49)		0.77 (0.39-1.54)	
Gastrointestinal symptoms						
No	123/1322	9.3	1.00	0.981	1.00	0.914
Yes	3/29	10.3	1.11 (0.38-3.29)		1.27 (0.41-3.91)	
Missing	37/393	9.4	1.01 (0.71-1.44)		0.99 (0.69-1.41)	

N=1744. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. For age, missing values are replaced with a constant value, here the median age, and an additional indicator variable is added to the model to indicate which values are missing (not shown). Global p-values by Wald test or test for trend ([a]). Each factor adjusted in separate model for *a priori* factors (age and clinic).

Appendix 3.14: Associations of behavioural factors with the detection of any bacterial enteric pathogen in HIV-negative MSM: sensitivity analyses incorporating worst- and best-case scenarios (low values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
PrEP use						
No	89/1197	7.4	1.00	<0.001	1.00	<0.001
Yes	74/547	13.5	1.82 (1.36-2.43)		1.76 (1.31-2.37)	
Interest in specific high-risk practices						
No	91/1184	7.7	1.00	<0.001	1.00	<0.001
Yes	72/560	12.9	1.67 (1.25-2.24)		1.65 (1.23-2.22)	
Number of sexual partners in last 3 months						
0-1	43/476	6.7	1.00	0.003	1.00	0.004
2-4	35/570	15.7	0.68 (0.44-1.04)	0.007[a]	0.67 (0.44-1.03)	0.017[a]
5-9	41/370	8.6	1.23 (0.82-1.84)		1.20 (0.80-1.80)	
10+	44/328	17.6	1.48 (1.00-2.21)		1.43 (0.96-2.13)	
Number of new sexual partners in last 3 months						
0-1	52/748	7.0	1.00	<0.001	1.00	<0.001
2-4	41/505	8.1	1.17 (0.79-1.73)	<0.001[a]	1.17 (0.79-1.73)	<0.001[a]
5-9	38/286	13.3	1.91 (1.29-2.84)		1.90 (1.28-2.83)	
10+	32/205	15.6	2.25 (1.49-3.39)		2.20 (1.45-3.34)	
Receptive anal sex in last 3 months						
No	20/207	9.7	1.00	0.868	1.00	0.706
Yes	143/1537	9.3	0.96 (0.62-1.50)		0.92 (0.58-1.45)	
Receptive oral sex in last 3 months						
No	22/196	11.2	1.00	0.334	1.00	0.191
Yes	141/1548	9.1	0.81 (0.53-1.24)		0.75 (0.49-1.16)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Last condomless sex						
Never/more than 6 weeks ago	58/705	8.2	1.00	0.027	1.00	0.032
Within 6 weeks	73/811	9.0	1.09 (0.79-1.52)		1.09 (0.78-1.51)	
Within 72 hours	32/228	14.0	1.71 (1.14-2.56)		1.69 (1.12-2.53)	
Gastrointestinal symptoms						
No	160/1715	9.3	1.00	0.852	1.00	0.676
Yes	3/29	10.3	1.11 (0.38-3.27)		1.27 (0.41-3.91)	

N=1744. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend ([a]). Each factor adjusted in separate model for *a priori* factors (age and clinic – not shown).

Appendix 3.15: Associations of behavioural factors with the detection of any bacterial enteric pathogen in HIV-negative MSM: sensitivity analyses incorporating worst- and best-case scenarios (high values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
PrEP use						
No	60/930	6.5	1.00	<0.001	1.00	<0.001
Yes	103/814	12.7	1.96 (1.45-2.66)		1.94 (1.43-2.64)	
Interest in specific high-risk practices						
No	72/941	7.7	1.00	0.009	1.00	0.012
Yes	91/803	11.3	1.48 (1.10-1.99)		1.47 (1.09-1.98)	
Number of sexual partners in last 3 months						
0-1	12/148	8.1	1.00	0.010	1.00	0.014
2-4	35/570	6.1	0.76 (0.40-1.42)	0.004[a]	0.72 (0.39-1.36)	0.025[a]
5-9	41/370	11.1	1.37 (0.74-2.53)		1.29 (0.70-2.38)	
10+	75/656	11.4	1.41 (0.79-2.53)		1.34 (0.75-2.39)	
Number of new sexual partners in last 3 months						
0-1	18/355	5.1	1.00	0.002	1.00	0.003
2-4	41/505	8.1	1.60 (0.94-2.74)	<0.001[a]	1.57 (0.92-2.68)	0.008[a]
5-9	38/286	13.3	2.62 (1.53-4.49)		2.56 (1.50-4.37)	
10+	66/598	11.0	2.18 (1.31-3.60)		2.14 (1.29-3.55)	
Receptive anal sex						
No	5/85	5.9	1.00	0.274	1.00	0.286
Yes	158/1659	9.5	1.62 (0.68-3.84)		1.59 (0.68-3.74)	
Receptive oral sex						
No	4/52	9.5	1.00	0.968	1.00	0.939
Yes	159/1702	9.3	0.98 (0.38-2.52)		0.96 (0.37-2.48)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Last condomless sex						
Never/more than 6 weeks ago	49/566	8.7	1.00	0.389	1.00	0.365
Within 6 weeks	73/811	9.0	1.04 (0.74-1.47)		1.04 (0.74-1.48)	
Within 72 hours	41/367	11.2	1.29 (0.87-1.91)		1.30 (0.88-1.94)	
Gastrointestinal symptoms						
No	123/1322	9.3	1.00	0.915	1.00	0.968
Yes	40/422	9.5	1.02 (0.73-1.43)		1.01 (0.72-1.42)	

N=1744. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend ([a]). Each factor adjusted in separate model for *a priori* factors (age and clinic – not shown).

Appendix 3.16: Associations of partner number with the detection of any bacterial enteric pathogen in <u>HIV-negative MSM</u>: sensitivity analyses using <u>single value imputation</u>

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Number of sexual partners in last 3 months						
0-1	12/148	8.1	1.00	0.007	1.00	0.012
2-4	66/898	7.4	0.91 (0.50-1.64)	0.001[a]	0.87 (0.48-1.57)	0.006[a]
5-9	41/370	11.1	1.37 (0.74-2.53)		1.30 (0.71-2.40)	
10+	44/328	13.4	1.65 (0.90-3.04)		1.56 (0.85-2.86)	
Number of sexual partners in last 3 months						
0-4	78/1046	7.5	1.00	0.001	1.00	0.002
5+	85/698	12.2	1.63 (1.22-2.19)		1.60 (1.19-2.14)	
Number of sexual partners in last 3 months						
Median (IQR)	4 (3-6)		1.02 (1.01-1.03)	<0.001	1.02 (1.01-1.03)	<0.001
Number of <u>new</u> sexual partners in last 3 months						
0-1	18/355	5.1	1.00	<0.001	1.00	<0.001
2-4	75/898	8.4	1.65 (1.00-2.71)	<0.001[a]	1.63 (0.99-2.68)	<0.001[a]
5-9	38/286	13.3	2.62 (1.53-4.49)		2.58 (1.51-4.40)	
10+	32/205	15.6	3.08 (1.77-5.34)		2.98 (1.72-5.16)	
Number of <u>new</u> sexual partners in last 3 months						
0-4	93/1253	7.4	1.00	<0.001	1.00	<0.001
5+	70/491	14.3	1.92 (1.43-2.57)		1.89 (1.41-2.55)	
Number of <u>new</u> sexual partners in last 3 months						
Median (IQR)	3 (2-5)		1.03 (1.01-1.04)	<0.001	1.02 (1.01-1.04)	<0.001

N=1744. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (Cis) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test or test for trend (a). Missing values replaced with median value. Adjusted Model: Each factor adjusted in separate model for a priori factors (age and clinic – not shown).

Appendix 3.17: Associations of socio-demographic, clinical and behavioural factors with the detection of any bacterial enteric pathogen in HIV-diagnosed MSM: sensitivity analysis using the missing indicator method

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Clinic						
DSE	34/248	13.7	1.00	0.122	1.00	0.145
56DS	10/124	8.1	0.59 (0.30-1.15)		0.60 (0.30-1.19)	
Age group						
16-34	20/154	13.0	1.00	0.546	1.00	0.609
35+	24/218	11.0	0.84 (0.48-1.47)		0.86 (0.49-1.52)	
Ethnic category						
White	33/272	12.1	1.00	0.861	1.00	0.899
Ethnic minority	10/86	11.6	0.96 (0.49-1.86)		0.96 (0.50-1.84)	
Missing	1/14	7.1	0.59 (0.09-4.01)		0.65 (0.10-4.26)	
Region of birth						
UK	19/151	12.6	1.00	0.865	1.00	0.879
Europe	14/104	13.5	1.07 (0.56-2.04)		1.01 (0.52-1.94)	
Rest of world	11/100	11.0	0.87 (0.44-1.76)		0.85 (0.43-1.70)	
Missing	0/17	0	omitted		omitted	
IMD quintile						
1-2 (Most deprived)	36/258	14.0	1.00	0.192	1.00	0.195
3	4/61	6.6	0.47 (0.17-1.27)		0.47 (0.17-1.25)	
4-5 (Least deprived)	4/52	7.7	0.55 (0.20-1.48)		0.58 (0.22-1.55)	
Missing	0/1	0	omitted		omitted	
Sexual orientation						
Gay	43/362	11.9	NA			
Bisexual/heterosexual	0/3	0				
Missing	1/7	14.3				

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Bacterial STI diagnosed at attendance						
No/unknown	29/275	10.6	1.00	0.196	1.00	0.209
Yes	15/97	15.5	1.47 (0.82-2.62)		1.46 (0.81-2.57)	
Bacterial STI diagnosed in last year						
No/unknown	18/173	10.4	1.00	0.431	1.00	0.683
Yes	26/199	13.1	1.26 (0.71-2.21)		1.13 (0.63-2.05)	
Interest in specific high-risk practices						
No	25/133	18.8	1.00	0.009	1.00	0.024
Yes	13/141	9.2	0.49 (0.26-0.92)		0.49 (0.26-0.91)	
Missing	6/98	6.1	0.33 (0.14-0.76)		0.38 (0.12-1.16)	
Number of sexual partners in last 3 months						
0-4	19/142	7.7	1.00	0.312	1.00	0.739
5+	19/144	12.4	0.99 (0.55-1.78)		0.99 (0.54-1.80)	
Missing	6/86	7.0	0.52 (0.22-1.26)		0.65 (0.22-1.98)	
Number of new sexual partners in last 3 months						
0-4	23/170	13.5	1.00	0.379	1.00	0.872
5+	13/101	12.9	0.95 (0.50-1.79)		0.96 (0.51-1.82)	
Missing	8/101	7.9	0.59 (0.27-1.26)		0.76 (0.28-2.09)	
Receptive anal sex in last 3 months						
No	2/7	28.6	1.00	0.076	1.00	0.130
Yes	37/279	13.3	0.46 (0.14-1.56)		0.40 (0.12-1.40)	
Missing	5/86	5.8	0.20 (0.05-0.87)		0.21 (0.05-0.96)	
Receptive oral sex in last 3 months						
No	2/5	40.0	1.00	0.020	1.00	0.069
Yes	37/279	13.3	0.33 (0.11-1.01)		0.33 (0.10-1.05)	
Missing	5/88	5.7	0.14 (0.04-0.56)		0.16 (0.03-0.77)	

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Last condomless sex						
Never/more than 6 weeks ago	5/55	9.1	1.00	0.181	1.00	0.368
Within 6 weeks	32/222	14.4	1.59 (0.65-3.89)		1.64 (0.66-4.07)	
Missing	7/95	7.4	0.81 (0.27-2.44)		1.02 (0.29-3.62)	
Gastrointestinal symptoms						
No	25/231	10.8	1.00	0.293	1.00	0.105
Yes	2/7	28.6	2.64 (0.77-9.03)		4.12 (1.07-15.9)	
Missing	17/134	12.7	1.17 (0.66-2.09)		0.86 (0.47-1.58)	

N=372. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. For age, missing values are replaced with a constant value, here the median age, and an additional indicator variable is added to the model to indicate which values are missing (not shown). Global p-values by Wald test. Each factor adjusted in separate model for *a priori* factors (age and clinic).

Appendix 3.18: Associations of behavioural factors with the detection of any bacterial enteric pathogen in HIV-diagnosed MSM: sensitivity analyses incorporating worst- and best-case scenarios (low values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Interest in specific high-risk practices						
No	31/231	13.4	1.00	0.231	1.00	0.098
Yes	13/141	9.2	0.69 (0.37-1.27)		0.57 (0.29-1.11)	
Number of sexual partners in last 3 months						
0-4	25/228	11.0	1.00	0.517	1.00	0.823
5+	19/144	13.2	1.20 (0.69-2.11)		1.07 (0.59-1.93)	
Number of new sexual partners in last 3 months						
0-4	31/271	11.4	1.00	0.703	1.00	0.972
5+	13/101	12.9	1.13 (0.61-2.06)		1.01 (0.55-1.88)	
Receptive anal sex in last 3 months						
No	7/93	7.5	1.00	0.152	1.00	0.503
Yes	37/279	13.3	1.76 (0.81-3.82)		1.40 (0.52-3.77)	
Receptive oral sex in last 3 months						
No	7/93	7.5	1.00	0.152	1.00	0.488
Yes	37/279	13.3	1.76 (0.81-3.82)		1.41 (0.53-3.72)	
Last condomless sex						
Never/more than 6 weeks ago	12/150	8.00	1.00	0.068	1.00	0.161
Within 6 weeks	32/222	14.4	1.80 (0.96-3.39)		1.63 (0.82-3.22)	
Gastrointestinal symptoms						
No	42/365	11.5	1.00	0.140	1.00	0.040
Yes	2/7	28.6	2.48 (0.74-8.30)		4.13 (1.07-16.0)	

N=372. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test. Each factor adjusted in separate model for *a priori* factors (age and clinic – not shown).

Appendix 3.19: Associations of behavioural factors with the detection of any bacterial enteric pathogen in HIV-diagnosed MSM: sensitivity analyses incorporating worst- and best-case scenarios (high values)

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Interest in specific high-risk practices						
No	25/133	18.8	1.00	0.003	1.00	0.007
Yes	19/239	8.0	0.42 (0.24-0.74)		0.45 (0.26-0.81)	
Number of sexual partners in last 3 months						
0-4	19/142	13.4	1.00	0.467	1.00	0.755
5+	25/230	10.9	0.81 (0.46-1.42)		0.91 (0.50-1.65)	
Number of new sexual partners in last 3 months						
0-4	23/170	13.5	1.00	0.353	1.00	0.754
5+	21/202	10.4	0.77 (0.44-1.34)		0.91 (0.49-1.67)	
Receptive anal sex in last 3 months						
No	2/7	28.6	1.00	0.140	1.00	0.088
Yes	42/365	11.5	0.40 (0.12-1.35)		0.34 (0.10-1.17)	
Receptive oral sex in last 3 months						
No	2/5	40.0	1.00	0.027	1.00	0.056
Yes	42/367	11.4	0.29 (0.09-0.87)		0.30 (0.09-1.03)	
Last condomless sex						
Never/more than 6 weeks ago	5/55	9.1	1.00	0.504	1.00	0.361
Within 6 weeks	39/317	12.3	1.35 (0.56-3.29)		1.53 (0.62-3.80)	
Gastrointestinal symptoms						
No	25/231	10.8	1.00	0.442	1.00	0.972
Yes	19/141	13.5	1.25 (0.71-2.18)		1.01 (0.54-1.90)	

N=372. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test. Each factor adjusted in separate model for *a priori* factors (age and clinic – not shown).

Appendix 3.20: Associations of partner number with the detection of any bacterial enteric pathogen in HIV-diagnosed MSM: sensitivity analyses using single value imputation

	n/N	Row %	Unadjusted PR (95% CI)	p-value	Adjusted PR (95% CI)	p-value
Number of sexual partners in last 3 months						
0-1	2/34	5.9	1.00	0.070	1.00	0.155
2-4	17/108	15.7	2.68 (0.65-11.0)	0.461[a]	2.77 (0.66-11.8)	0.647[a]
5-9	12/156	7.7	1.31 (0.31-5.59)		1.53 (0.32-7.37)	
10+	13/74	17.6	2.99 (0.71-12.5)		3.01 (0.71-12.8)	
Number of sexual partners in last 3 months						
0-4	19/142	13.4	1.00	0.467	1.00	0.755
5+	25/230	10.9	0.81 (0.46-1.42)		0.91 (0.50-1.65)	
Number of sexual partners in last 3 months						
Median (IQR)	5 (3-7)		1.01 (1.00-1.03)	0.123	1.01 (0.99-1.03)	0.227
Number of new sexual partners in last 3 months						
0-1	10/87	11.5	1.00	0.931	1.00	0.952
2-4	21/184	11.4	0.99 (0.49-2.02)	0.461[a]	1.18 (0.56-2.52)	0.441[a]
5-9	7/61	11.5	1.00 (0.40-2.48)		1.03 (0.41-2.61)	
10+	6/40	15.0	1.31 (0.51-3.35)		1.25 (0.49-3.20)	
Number of new sexual partners in last 3 months						
0-4	31/271	11.4	1.00	0.878	1.00	0.972
5+	13/101	12.9	1.13 (0.61-2.06)		1.01 (0.55-1.87)	
Number of new sexual partners in last 3 months						
Median (IQR)	3 (2-5)		1.01 (0.99-1.03)	0.488	1.00 (0.98-1.03)	0.694

N=372. Unadjusted and adjusted prevalence ratios (PRs) and 95% confidence intervals (CIs) calculated using modified Poisson regression with robust error variance. Global p-values by Wald test. Missing values for partner number replaced with median value. Each factor adjusted in separate model for *a priori* factors (age and clinic – not shown).

Appendix for Chapter 4

Appendix 4.1: Standardised shigellosis exposure questionnaire

Questionnaire contains public sector information licensed under the Open Government Licence v3.0

Shigella *flexneri/dysenteriae/boydii* questionnaire

1: QUESTIONNAIRE DETAILS

Interviewer name:		Interview date :	
Interviewer organisation:		Interviewer Telephone:	
Person interviewed name and relationship to case:			

2: CASE CLASSIFICATION AND ID

PHE reference no. e.g. HPZone					
Environmental health ID no.(if relevant)					
If relevant: GUM clinic name		GUM clinician		GUM clinic no.	
Local laboratory result: Culture					
PCR					
Local laboratory specimen number		Local Laboratory name			
Reference laboratory result					

		Confirmed	Probable	Possible
Classification	*Shigella flexneri*	☐	☐	☐
	Shigella dysenteriae	☐	☐	☐
	Shigella boydii	☐	☐	☐
	Shigella species	☐	☐	☐

3: PERSONAL AND IDENTIFYING DETAILS

First name:			Family name/Surname:		
Address:					
Postcode:		Tel (h):		Tel (m):	
Email:					
Sex:	M ☐ F ☐	Date of birth (dd/mm/yyyy):		Age:	yrs
NHS No:		GP name:			
GP address:				GP Tel:	
Are there any children living in the household? (other than the case)		Y ☐	N ☐	How many? ☐	

Public Health England is committed to ensuring all individuals are treated equally and fairly, and therefore we ask all people the following questions about their ethnicity and sexual orientation. This also helps us to identify and investigate outbreaks.

Ethnicity:				
White	☐ British	☐ Irish		☐ Other (please state)
Mixed	☐ White/Black Caribbean	☐ White/Black African	☐ White/Asian	☐ Other (please state)
Asian/Asian British	☐ Indian	☐ Pakistani	☐ Bangladeshi	☐ Other (please state)
Black/Black British	☐ Caribbean	☐ African		☐ Other (please state)
Chinese	☐ Chinese			
Other	☐ (please state)			

Sexual identity (if over 18 years old)
I will now read out a list of terms people sometimes use to describe how they think of themselves.
☐ Heterosexual or straight
☐ Gay or lesbian
☐ Bisexual
☐ Other
As I read the list again please say 'yes' when you hear the option that best describes how you think of yourself.(Pause briefly after each option during second reading).
☐ Spontaneous don't know/refusal

4: RISK GROUPS

Occupation			
Work/School/Nursery name:		Tel:	
Contact person			
Address:			
Postcode:		Date last attended:	

Does the patient fit into any of the following categories (tick all that apply)?:		
Group A	☐	Any person of doubtful personal hygiene or with unsatisfactory toilet, hand washing or hand drying facilities at home, work or school. Particular consideration should be given as to whether individual infant-school-aged children (aged 6 or 7 years) are able to satisfactorily observe good personal hygiene. Health protection personnel (LA and HPU) should agree locally on how to make this assessment in engagement with parents or teachers/carers.
Group B	☐	All children aged five years old or under including those who attend school, pre-school, nursery or other childcare or minding groups.
Group C	☐	People whose work involves preparing or serving unwrapped food to be served raw or not subjected to further heating.
Group D	☐	Clinical, social care or nursery staff who work with young children, the elderly, or other particularly vulnerable people, and whose activities increase the risk of transferring infection via the faecal-oral route. Such activities include helping with feeding or handling objects that could be transferred to the mouth
No risk group	☐	

5: CLINICAL DETAILS

Onset date:		Still ill:		If no, →	Duration of illness (days):	
Symptoms:	Diarrhoea Yes ☐ No ☐			Vomiting Yes ☐ No ☐		
	Fever Yes ☐ No ☐			Blood in stools Yes ☐ No ☐		
	Abdominal pain Yes ☐ No ☐			Mucus in stool Yes ☐ No ☐		
Healthcare sought from:	☐ GP visit	☐ A&E	☐ Sexual health clinic	☐ Other (specify):		
Date of stool sample:						
Admitted to hospital for this illness:	Yes ☐ No ☐			Admission date:		
Hospital name:				Duration of stay (d):		
Treated with antibiotics:	Yes ☐ No ☐			If Y, specify:		
Has the patient heard of *Shigella* before?	Yes ☐ No ☐					

6: TRAVEL

6.1 In the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* days prior to illness, did you arrive or return to the UK from **ABROAD**?			Yes ☐ No ☐
Specify countries visited (from most recent)			
Country/Region	**Date arrived**	**Date departed**	**Details**
6.1 In the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* days prior to illness, did you travel elsewhere **WITHIN** the UK?			
Town/Resort	**Date arrived**	**Date departed**	**Details**

7: OTHER RISK FACTORS

7.1 In the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* days prior to illness, did you have any contact with:			
Anyone with diarrhoea?	Yes ☐ No ☐	If yes, details:	
Children under 5 years?	Yes ☐ No ☐	If yes, details:	
Visitors from UK or overseas?	Yes ☐ No ☐	If yes, details:	
Attend a gym/ swimming pool/other communal sports facilities?	Yes ☐ No ☐	If yes, where?	

8: SEXUAL CONTACT **(ONLY for MALE cases aged18 years or older. If under 18 years please go straight to section 9). Please ask about sexual contact regardless of travel history or other identified risk factors.**

8.1 We ask all adult men with Shigella about sexual contact as Shigella can be sexually transmitted via the faecal-oral route. Men who have sex with men (MSM) are at greater risk of acquiring the infection than others and there have been outbreaks of shigella among the MSM population.

Did you have sexual contact with anyone in the **FOUR** (*boydii, flexneri)* or **SEVEN** (*dysenteriae)* days prior to illness			Yes ☐ No ☐
IF yes, was this with a	Male ☐ and/or	Female ☐	Prefer not to answer ☐

8.2 Public Health Advice: If the case is a MSM - please remember to provide them information on how to reduce transmission as per PHE leaflet (link below) and recommend they attend their local sexual health clinic for STI screening.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/323532/Shigella_leaflet.pdf

If the case has been abroad to areas with a high risk for shigella infection in the FOUR *(boydii, flexneri)* or SEVEN (*dysenteriae)* days prior to illness, AND/OR sexual transmission has been identified as a likely source, please go straight to section 10.

High risk area - South America, Asia (including the Middle East) and Africa: Low risk areas - Europe, North America and Australia:

9: FOOD AND WATER HISTORY

9.1 Food prepared at home

Please list all food eaten in the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* before you became unwell. Describe what was eaten and when in the appropriate section on the chart. If possible give details of the shop where it was bought. **Routine cooking kills *Shigella* bacteria so particular attention should be given to raw and uncooked food, as well as baby foods.**

Days pre-onset	Date	Breakfast	Lunch	Evening meal	Snacks
1					
2					
3					
4					
5					
6					
7					

9.2 In the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* days before you became unwell, did you eat any food that was not prepared at home, either in this country or abroad (e.g. hotels, restaurants, cafes, pubs; school and work canteens; takeaways, fast food outlets; barbecues and picnics; social events; other people's homes)? If yes, enter details below

Date	Description of food	Establishment where food obtained

9.3 In the **FOUR** *(boydii, flexneri)* or **SEVEN** (*dysenteriae)* days before you became unwell, what was the source of your drinking water? (tick all that apply)

Mains ☐	Private ☐	Bottled ☐
Filtered ☐	Well ☐	Spring ☐
Other (specify) ☐		

10: PLEASE COMPLETE CONTACT SHEET ON THE FOLLOWING PAGE

11: FURTHER INFORMATION

Is this case part of an outbreak?	Yes ☐ No ☐	**Setting**
Are there any other possible exposures to infection not already discussed		
May we contact you again if we need to ask any further questions?	Yes ☐	No ☐

Please remember for provide public health advice to reduce risk of ongoing transmission

Appendix 4.2: Information governance and Public Health England approval letters

Public Health England

PHE Research Support and Governance Office

8th June 2017

Gwenda Hughes
PHE

Dear Gwenda

PROJECT TITLE: Epidemiology of sexually transmissible enteric infections and their capacity for acquiring antimicrobial resistance genes in sexual networks of men who have sex with men

Approval date: 8th June 2017

Thank you for submitting your project to the PHE Research Support and Governance Office (RSGO).

Following an internal review it has been decided that this project is not research, but is a surveillance study. This decision has been made based on the protocol provided on 16th May 2017.

Based on this information the RSGO approves this investigation. This approval is given on the understanding that the surveillance

- Is conducted in accordance with your protocol version 1.0 dated 25th April 2017
- Is compliant with all relevant regulatory requirements
- Has sign off from the Caldicott Panel.

If any aspect of the study should change you will need to inform this office to ensure this approval remains in place.

If you need any further support or information, please do not hesitate to contact the PHE RSGO.

Wishing you every success with your surveillance study

Yours sincerely,

Elizabeth Coates
Head of Research Governance
Public Health England

cc: Holly Mitchell

Public Health England

Protecting and improving the nation's health

NIS Caldicott Review Panel

Date 15/06/17
Holly Mitchell

Caldicott Opinion

Dear Holly Mitchell,

Investigation into the epidemiology of gastrointestinal infections associated with sexual transmission in MSM

You are seeking clarification on whether you are covered by the current PHE Section 251 of the NHS Act 2006 and the Health Service (Control of Patient Information) Regulations 2002 ('section 251 support') approval we hold for our surveillance systems, to use pseudonymised surveillance data for your analysis of the epidemiological factors associated with the transmission of shigella in the MSM community. The proposed analysis is required better understand the emerging trend of sexually transmitted enteric infections. The overall aim of the analysis is to better understand the epidemiology of sexually transmissible enteric infections so that improved public health and clinical control measures can be developed and tested.

In summary, the study is an analysis of a linked dataset which consists of epidemiological and behavioural data collected as part of the enhanced surveillance for shigella, a notifiable infection, routine HIV surveillance and routine STI surveillance. This data will be linked with the existing phylogenetic data collected through the reference lab as part of their surveillance of shigella.

You will use molecular typing data from positive Shigella spp. specimens sent to the GBRU (accessible via Gastro Data Warehouse). You will link this data to epidemiological data collected via patient questionnaires to create an enhanced dataset which will enable the study of phylogenetic clusters and associated risk factors to a far greater resolution than has been previously studied. These data will be directly linked using common identifiers including NHS number.

You will link GDW data to HARS to identify enteric infections among HIV positive individuals and to explore the changing rate of enteric infections over time among these individuals. A standard PHE protocol will be used to create an irreversibly unlinked anonymous dataset.

To facilitate this work, you propose a direct and probabilistic linkage process to be managed within PHE secure servers which will be controlled under the existing SLSPs of the information assets involved. The final dataset will be stripped of identifiers but will not achieve the ICO standard of anonymisation.

You have confirmed that only named PHE staff will be allowed to access these data and that data access will be restricted to the minimum number of staff members necessary to process and use the data. We request that linkage with identifiable data be performed by existing PHE staff to minimise the total number of people with access to the identifiable data.

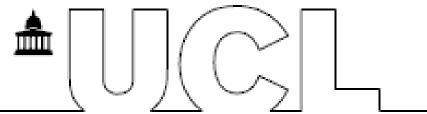
Based on the information provided above, I can confirm that the proposed project, linkage and affiliated activities such as analysis for peer reviewed publication is covered under the current PHE permissions for public health surveillance under Section 251 of the NHS Act 2006 and the Health Service (Control of Patient Information) Regulations 2002 ('section 251 support') and that the resultant data base for analysis will be using only pseudonymised data, with risk of deductive disclosure. This data will be retained within PHE and deleted at the end of the project.

If there are any changes to the data being accessed which would result in further identifiable information being required please do seek further input from the Caldicott team.

Kind regards,

Dr Kevin Dunbar

Associate Caldicott guardian, CIDSC



10th August 2020

To whom it may concern,

Holly Mitchell is a PhD candidate at UCL who is funded by the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Blood Borne and Sexually Transmitted Infections (BBSTI). Her primary supervisor is Professor Gwenda Hughes, Head of Blood Safety, Hepatitis, STIs and HIV in the National Infection Service at Public Health England, and secondary supervisor is Dr Nigel Field, Associate Professor at UCL Institute for Global Health.

Holly's work has two major components:

- A clinical research study was undertaken through an NHS sexual health clinic with ethical approval granted by the London Harrow NHS Research Ethics Committee and NHS Health Research Authority (17/LO/1722) and confirmed by the Chelsea and Westminster Hospital NHS Foundation Trust Research and Development Office (C&W17/056).
- A second study used clinical data collected by Public Health England (PHE), which was granted authority to handle patient data for public health monitoring and infection control under Section 251 of the UK NHS Act of 2006. PHE has the power of approval for public health surveillance activity and operates a PHE Caldicott Panel. The PHE Research Support and Governance Office approved the analyses undertaken within the PhD as not requiring ethical approval. In accordance with standard PHE procedures, the analyses were then reviewed and approved by the PHE Caldicott Panel in June 2017. Several further steps were taken to provide governance and data security; Holly held an honorary contract with PHE to undertake the work, the final dataset was irreversibly anonymised prior to analysis, and the work was performed on PHE computers within the PHE firewall. This approach mirrored that taken for all other work funded by the HPRU BBSTI using PHE clinical data and undertaken at UCL.

Dr Nigel Field

Director, Centre for Molecular Epidemiology & Translational Research
Institute for Global Health
University College London

Professor Gwenda Hughes

Head of Blood Safety, Hepatitis, STIs & HIV and Deputy Director
National Infection Service
Public Health England

Appendix 4.3: Summary of HIV matching algorithm and number of matched *S. flexneri* isolates at each hierarchy level

Match type	Match description	N (%)
1a	sdex, dob, initial, sex, postcode	34 (19.7)
1aa	sdex, dob, initial, sex, postal district	4 (2.3)
1b	sdex, dob, initial, sex, lares	75 (43.4)
1c	sdex, dob, initial, sex	45 (26.0)
2a	sdex, dob, sex, postcode	0
2aa	sdex, dob, sex, postal district	0
2b	sdex, dob, sex, lares	3 (1.7)
2c	sdex, dob, sex, phecentreres	0
3a	sdex letter, dob, initial, sex, postcode	0
3aa	sdex letter, dob, initial, sex, postal district	0
3b	sdex letter, dob, initial, sex, lares	3 (1.7)
4a	sdex, sex, initial, postcode, day of birth, month of birth	0
4aa	sdex, sex, initial, postal district, day of birth, month of birth	0
4b	sdex, sex, initial, postcode, month of birth, year of birth	1 (0.6)
4bb	sdex, sex, initial, postal district, month of birth, year of birth	0
4c	sdex, sex, initial, postcode, day of birth, year of birth	0
4cc	sdex, sex, initial, postal district, day of birth, year of birth	0
5a	sdex, sex, initial, lsoa, day of birth, month of birth	0
5b	sdex, sex, initial, lsoa, month of birth, year of birth	1 (0.6)
5c	sdex, sex, initial, lsoa, day of birth, year of birth	0
6a	sdex, sex, initial, lares, day of birth, month of birth	0
6b	sdex, sex, initial, lares, month of birth, year of birth	3 (1.7)
6c	sdex, sex, initial, lares, day of birth, year of birth	0
7	dob, sex, postcode, sdex and initial swapped	2 (1.2)
8	sex, initial, dob, f3sdex, phecentreres	2 (1.2)
Total		**173**

Sdex: Soundex code, dob: date of birth, lares: local authority of residence, lsoa: lower layer super output area of residence, f3sdex: first three digits of Soundex code, phecentreres: PHE centre of residence

Appendix 4.4: Characteristics associated with composite measures of clinical symptoms or outcomes in adults diagnosed with *S. flexneri* and with linked questionnaire data

	Clinical symptoms: blood and/or mucus in stools			Clinical outcomes: hospital admission and/or antimicrobial use		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group (n=156)						
Non MSM	34/61 (55.7)	1.00	1.00	35/61 (57.4)	1.00	1.00
MSM	68/95 (71.6)	2.00 (1.02-3.92)	1.88 (0.95-3.73)	76/95 (80.0)	2.97 (1.45-6.07)	3.07 (1.49-6.31)
p-value		0.043	0.070		0.003	0.002
HIV status (n=164)						
Negative/unknown	71/112 (63.4)	1.00	1.00	74/112 (66.1)	1.00	1.00
Living with HIV	36/52 (69.2)	1.30 (0.64-2.63)	1.26 (0.62-2.57)	43/52 (82.7)	2.45 (1.08-5.56)	2.50 (1.08-5.57)
p-value		0.463	0.516		0.024	0.024
Serotype (n=152)						
2a	70/105 (66.7)	1.00	1.00	76/105 (72.4)	1.00	1.00
Other	29/47 (61.7)	0.81 (0.39-1.65)	0.80 (0.39-1.65)	31/47 (66.0)	0.74 (0.35-1.55)	0.74 (0.35-1.55)
p-value		0.554	0.545		0.426	0.426
Lineage (n=165)						
MSM clade	69/97 (71.1)	1.00	1.00	75/97 (77.3)	1.00	1.00
Travel-associated lineage	39/68 (57.4)	0.55 (0.28-1.05)	0.60 (0.31-1.16)	43/68 (63.2)	0.50 (0.25-1.00)	0.49 (0.25-0.99)
p-value		0.068	0.131		0.050	0.045
Azithromycin resistance (n=165)						
No	44/74 (59.5)	1.00	1.00	45/74 (60.8)	1.00	1.00
Yes	64/91 (70.3)	1.62 (0.85-3.08)	1.43 (0.73-2.79)	73/91 (80.2)	2.61 (1.30-5.24)	2.75 (1.35-5.61)
p-value		0.145	0.294		0.006	0.005

	Clinical symptoms: blood and/or mucus in stools			Clinical outcomes: hospital admission and/or antimicrobial use		
Foreign travel (n=165)						
No	71/104 (68.3)	1.00	1.00	83/104 (79.8)	1.00	1.00
Yes	37/61 (60.7)	0.72 (0.37-1.39)	0.77 (0.39-1.50)	35/61 (57.4)	0.34 (0.17-0.68)	0.34 (0.17-0.68)
p-value		0.323	0.436		0.002	0.002
Age group (n=165)						
18-24	8/15 (53.3)	0.69 (0.23-2.07)	0.69 (0.23-2.07)	10/15 (66.7)	0.74 (0.23-2.33)	0.74 (0.23-2.33)
25-34	42/72 (73.7)	1.69 (0.82-3.48)	1.69 (0.82-3.48)	40/72 (70.2)	0.87 (0.42-1.79)	0.87 (0.42-1.79)
≥35	58/93 (62.4)	1.00	1.00	68/93 (73.1)	1.00	1.00
p-value		0.212	0.212		0.845	0.845
Per year (age as a continuous variable)		0.98 (0.95-1.00)	0.98 (0.95-1.00)		1.00 (0.98-1.03)	1.00 (0.98-1.03)
p-value (age as a continuous variable)		0.047	0.047		0.959	0.959
Ethnic group (n=147)						
White	72/97 (74.2)	1.00	1.00	77/97 (79.4)	1.00	1.00
Ethnic minority	28/50 (56.0)	0.44 (0.22-0.91)	0.46 (0.22-0.95)	30/50 (60.0)	0.39 (0.18-0.82)	0.39 (0.18-0.82)
p-value		0.026	0.036		0.014	0.013
IMD quintile (n=164)						
1-2 (most deprived)	84/125 (67.2)	1.00	1.00	90/125 (72.0)	1.00	1.00
3	13/23 (56.5)	0.63 (0.26-1.57)	0.71 (0.28-1.80)	17/23 (73.9)	1.10 (0.40-3.02)	1.09 (0.39-3.01)
4-5 (least deprived)	11/16 (68.8)	1.07 (0.35-3.29)	1.22 (0.39-3.81)	11/16 (68.8)	0.86 (0.28-2.64)	0.84 (0.27-2.62)
p-value		0.600	0.700		0.940	0.938

Total numbers vary for each question due to missing items. IMD=Index of Multiple Deprivation. Unadjusted and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age only as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over.

Appendix 4.5: Characteristics associated with blood in stools, mucus in stools and fever in adults diagnosed with *S. flexneri* and with linked questionnaire data: <u>sensitivity analysis</u>

	Blood in stools (n/N=90/148)			Mucus in stools (n/N=60/128)			Fever (n/N=107/146)		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group									
Non-MSM	26/55 (47.3)	1.00	1.00	14/48 (29.2)	1.00	1.00	32/52 (61.5)	1.00	1.00
MSM	60/85 (70.6)	2..68 (1.32-5.42)	2.66 (1.30-5.44)	41/72 (56.9)	3.21 (1.48-6.99)	3.20 (1.46-7.03)	70/87 (80.5)	2.57 (1.19-5.56)	2.67 (1.22-5.82)
p-value		0.006	0.007		0.003	0.003		0.016	0.013
HIV status									
Negative/unknown	59/100 (59.0)	1.00	1.00	33/86 (38.4)	1.00	1.00	68/98 (69.4)	1.00	1.00
Living with HIV	30/47 (63.8)	1.23 (0.60-2.51)	1.25 (0.61-2.58)	26/41 (63.4)	2.78 (1.29-6.01)	2.79 (1.29-6.04)	38/47 (80.9)	1.86 (0.80-4.33)	1.86 (0.80-4.32)
p-value		0.575	0.542		0.008	0.008		0.137	0.139
Serotype									
2a	60/90 (66.7)	1.00	1.00	38/73 (52.1)	1.00	1.00	72/92 (78.3)	1.00	1.00
Other	23/45 (51.1)	0.52 (0.25-1.09)	0.52 (0.25-1.09)	16/43 (37.2)	0.55 (0.25-1.18)	0.55 (0.26=1.20)	27/43 (62.8)	0.47 (0.21-1.04)	0.46 (0.21-1.02)
p-value		0.082	0.084		0.120	0.131		0.062	0.058
Lineage									
MSM clade	58/87 (66.7)	1.00	1.00	45/75 (60.0)	1.00	1.00	70/87 (80.5)	1.00	1.00
Travel-associated lineage	32/61 (52.5)	0.55 (0.28-1.08)	0.58 (0.30-1.15)	15/53 (28.3)	0.26 (0.12-0.56)	0.25 (0.12-0.55)	37/59 (62.7)	0.41 (0.19-0.86)	0.38 (0.18-0.82)
p-value		0.082	0.121		<0.001	<0.001		0.018	0.013
Azithromycin resistance									
No	34/65 (52.3)	1.00	1.00	19/57 (33.3)	1.00	1.00	44/64 (68.8)	1.00	1.00
Yes	56/83 (67.5)	1.89 (0.97-3.69)	1.77 (0.90-3.49)	41/71 (57.8)	2.73 (1.32-5.64)	2.76 (1.32-5.77)	63/82 (76.8)	1.51 (0.72-3.15)	1.58 (0.75-3.35)
p-value		0.061	0.099		0.006	0.006		0.275	0.229
Foreign travel									
No	60/91 (65.9)	1.00	1.00	44/75 (58.7)	1.00	1.00	72/91 (79.1)	1.00	1.00
Yes	30/57 (52.6)	0.57 (0.29-1.13)	0.60 (0.30-1.20)	16/53 (30.2)	0.30 (0.14-0.64)	0.30 (0.14-0.64)	35/55 (63.6)	0.46 (0.22-0.97)	0.44 (0.21-0.94)
p-value		0.108	0.149		0.001	0.001		0.043	0.033

	Blood in stools (n/N=90/148)			Mucus in stools (n/N=60/128)			Fever (n/N=107/146)		
Age group									
18-24	8/14 (57.1)	0.99 (0.31-3.10)	N/A	N/A	N/A	N/A	6/11 (54.6)	0.39 (0.11-1.43)	N/A
25-34	36/54 (66.7)	1.48 (0.72-3.03)		26/59 (44.1)	0.81 (0.40-1.63)		40/54 (74.1)	0.94 (0.42-2.07)	
≥35	46/80 (57.5)	1.00		34/69 (49.3)	1.00		61/81 (75.3)	1.00	
p-value		0.539			0.556			0.375	
Per year (continuous variable)		0.98 (0.95-1.00)			1.00 (0.97-1.02)			1.01 (0.98-1.04)	
p-value (continuous variable)		0.076			0.761			0.628	
Ethnic group									
White	59/90 (65.6)	1.00	1.00	46/77 (59.7)	1.00	1.00	66/87 (75.9)	1.00	1.00
Ethnic minority	24/44 (54.6)	0.63 (0.30-1.32)	0.61 (0.29-1.30)	12/40 (30.0)	0.29 (0.13-0.65)	0.29 (0.13-0.65)	31/46 (67.4)	0.66 (0.30-1.45)	0.66 (0.30-1.46)
p-value		0.220	0.202		0.002	0.002		0.300	0.311
IMD quintile									
1-2 (most deprived)	70/114 (61.4)	1.00	1.00	47/96 (49.0)	1.00	1.00	84/112 (75.0)	1.00	1.00
3	10/21 (47.6)	0.57 (0.22-1.46)	0.64 (0.24-1.66)	9/21 (42.9)	0.79 (0.30-2.03)	0.79 (0.30-2.07)	18/23 (78.3)	1.20 (0.41-3.54)	1.19 (0.37-3.34)
4-5 (least deprived)	10/13 (76.9)	2.10 (0.55-8.03)	2.50 (0.63-9.87)	4/11 (36.4)	0.60 (0.16-2.17)	0.60 (0.16-2.24)	5/11 (45.5)	0.28 (0.08-0.98)	0.25 (0.07-0.91)
p-value		0.220	0.203		0.670	0.698		0.119	0.095

Total numbers vary for each question due to missing items. IMD=Index of Multiple Deprivation. Unadjusted and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age only as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over.

Appendix 4.6: Characteristics associated with hospital admission and antimicrobial use in adults diagnosed with *S. flexneri* and with linked questionnaire data: sensitivity analysis

	Hospital admission (n/N=48/155)			Antimicrobial use (n/N=113/157)		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group						
Non MSM	12/58 (20.7)	1.00	1.00	32/57 (56.1)	1.00	1.00
MSM	33/88 (37.5)	2.30 (1.07-4.96)	2.35 (1.08-5.09)	74/92 (80.4)	3.21 (1.54-6.69)	3.36 (1.60-7.07)
p-value		0.029	0.026		0.002	0.001
HIV status						
Negative/unknown	30/108 (27.8)	1.00	1.00	71/107 (66.4)	1.00	1.00
Living with HIV	17/46 (37.0)	1.52 (0.73-3.17)	1.52 (0.73-3.17)	41/49 (83.7)	2.60 (1.10-6.12)	2.62 (1.11-6.18)
p-value		0.262	0.262		0.021	0.020
Serotype						
2a	36/96 (37.5)	1.00	1.00	72/98 (73.5)	1.00	1.00
Other	6/47 (12.8)	0.24 (0.09-0.63)	0.24 (0.09-0.63)	30/46 (65.2)	0.68 (0.32-1.44)	0.67 (0.32-1.43)
p-value		0.001	0.001		0.314	0.308
Lineage						
MSM clade	31/88 (35.2)	1.00	1.00	73/92 (79.4)	1.00	1.00
Travel-associated lineage	17/67 (25.4)	0.63 (0.31-1.26)	0.62 (0.30-1.27)	40/65 (61.5)	0.42 (0.20-0.85)	0.39 (0.19-0.80)
p-value		0.186	0.186		0.015	0.010
Azithromycin resistance						
No	21/72 (29.2)	1.00	1.00	42/69 (60.9)	1.00	1.00
Yes	27/83 (32.5)	1.17 (0.59-2.32)	1.17 (0.58-2.35)	71/88 (80.7)	2.68 (1.31-5.50)	2.92 (1.40-6.11)
p-value		0.651	0.660		0.006	0.004

	Hospital admission (n/N=48/155)			Antimicrobial use (n/N=113/157)		
Foreign travel						
No/unknown	38/96 (39.6)	1.00	1.00	80/99 (80.8)	1.00	1.00
Yes	10/59 (17.0)	0.31 (0.14-0.69)	0.31 (0.14-0.68)	33/58 (56.9)	0.31 (0.15-0.65)	0.30 (0.14-0.62)
p-value		0.002	0.002		0.001	0.001
Age group						
18-24	3/15 (28.4)	0.63 (0.16-2.42)	N/A	10/14 (71.4)	0.88 (0.25-3.10)	N/A
25-34	20/52 (38.5)	1.58 (0.76-3.25)		38/55 (69.1)	0.79 (0.38-1.66)	
≥35	25/88 (28.4)	1.00		65/88 (73.5)	1.00	
p-value		0.286			0.826	
Per year (continuous variable)		1.00 (0.97-1.02)			1.01 (0.98-1.03)	
p-value (continuous variable)		0.914			0.667	
Ethnic group						
White	26/91 (28.6)	1.00	1.00	77/96 (80.2)	1.00	1.00
Ethnic minority	17/48 (35.4)	1.37 (0.65-2.89)	1.64 (0.75-3.60)	25/48 (52.1)	0.27 (0.13-0.57)	0.26 (0.12-0.56)
p-value		0.409	0.215		<0.001	<0.001
IMD quintile						
1-2 (Most deprived)	37/117 (31.6)	1.00	1.00	85/120 (70.8)	1.00	1.00
3	8/22 (36.4)	1.24 (0.48-3.20)	1.23 (0.47-3.22)	17/22 (77.3)	1.13 (0.34-3.80)	1.10 (0.32-3.73)
4-5 (Least deprived)	3/15 (20.0)	0.54 (0.14-2.03)	0.54 (0.14-2.05)	11/15 (73.3)	1.40 (0.48-4.09)	1.37 (0.46-4.03)
p-value		0.541	0.542		0.815	0.843

Total numbers vary for each question due to missing items. IMD=Index of Multiple Deprivation. Unadjusted and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age only as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over.

Appendix 4.7: Characteristics associated with composite measures of clinical symptoms or outcomes in adults diagnosed with *S. flexneri* and with linked questionnaire data: sensitivity analysis

	Clinical symptoms: blood and/or mucus in stools (n/N=108/145)			Clinical outcomes: hospital admission and/or antimicrobial use (n/N=118/155)		
	n/N (%)	**OR (95% CI)**	**aOR (95% CI)**	**n/N (%)**	**OR (95% CI)**	**aOR (95% CI)**
Exposure group						
Non MSM	34/52 (65.4)	1.00	1.00	35/57 (61.4)	1.00	1.00
MSM	68/85 (80.0)	2.12 (0.97-4.62)	2.07 (0.94-4.53)	76/90 (84.4)	3.41 (1.56-7.45)	3.55 (1.62-7.82)
p-value		0.059	0.070		0.002	0.001
HIV status						
Negative/unknown	71/97 (73.2)	1.00	1.00	74/107 (69.2)	1.00	1.00
Living with HIV	36/47 (76.6)	1.20 (0.53-2.70)	1.21 (0.53-2.72)	43/47 (91.5)	4.79 (1.59-14.5)	4.82 (1.60-14.5)
p-value		0.660	0.651		0.001	0.001
Serotype						
2a	70/89 (78.7)	1.00	1.00	76/96 (79.2)	1.00	1.00
Other	29/43 (67.4)	0.56 (0.25-1.27)	0.56 (0.25-1.28)	31/46 (67.4)	0.54 (0.25-1.20)	0.54 (0.25-1.19)
p-value		0.169	0.174		0.133	0.131
Lineage						
MSM clade	69/86 (80.2)	1.00	1.00	75/90 (83.3)	1.00	1.00
Travel-associated lineage	39/59 (66.1)	0.48 (0.23-1.02)	0.50 (0.23-1.07)	43/65 (66.2)	0.39 (0.18-0.83)	0.37 (0.17-0.79)
p-value		0.057	0.075		0.014	0.010
Azithromycin resistance						
No	44/63 (69.8)	1.00	1.00	45/69 (65.2)	1.00	1.00
Yes	64/82 (78.1)	1.54 (0.73-3.25)	1.46 (0.68-3.12)	73/86 (84.9)	2.99 (1.39-6.47)	3.24 (1.47-7.14)
p-value		0.263	0.332		0.004	0.003

	Clinical symptoms: blood and/or mucus in stools (n/N=108/145)			Clinical outcomes: hospital admission and/or antimicrobial use (n/N=118/155)		
Foreign travel						
No	71/90 (78.9)	1.00	1.00	83/98 (84.7)	1.00	1.00
Yes	37/55 (67.3)	0.55 (0.26-1.17)	0.57 (0.27-1.22)	35/57 (61.4)	0.29 (0.13-0.62)	0.27 (0.12-0.59)
p-value		0.123	0.151		0.001	<0.001
Age group						
18-24	8/12 (66.7)	0.72 (0.20-2.66)	N/A	10/14 (71.4)	0.74 (0.21-2.60)	N/A
25-34	42/54 (77.8)	1.27 (0.56-2.86)		40/53 (75.5)	0.90 (0.41-2.01)	
≥35	58/79 (73.4)	1.00		68/88 (77.3)	1.00	
p-value		0.695			0.887	
Per year (age as a continuous variable)		0.98 (0.96-1.01)			1.01 (0.98-1.04)	
p-value (age as a continuous variable)		0.274			0.698	
Ethnic group						
White	72/88 (81.8)	1.00	1.00	77/85 (81.1)	1.00	1.00
Ethnic minority	28/43 (65.1)	0.42 (0.18-0.95)	0.41 (0.18-0.94)	30/47 (63.8)	0.41 (0.19-0.91)	0.41 (0.18-0.89)
p-value		0.038	0.035		0.028	0.025
IMD quintile						
1-2 (most deprived)	84/111 (75.7)	1.00	1.00	90/118 (76.3)	1.00	1.00
3	13/21 (61.9)	0.52 (0.20-1.39)	0.56 (0.21-1.52)	17/22 (77.3)	1.06 (0.36-3.13)	1.03 (0.35-3.06)
4-5 (least deprived)	11/13 (84.6)	1.77 (0.37-8.48)	1.96 (0.40-9.54)	11/15 (73.3)	0.86 (0.25-2.90)	0.82 (0.24-2.83)
p-value		0.291	0.306		0.961	0.950

Total numbers vary for each question due to missing items. IMD=Index of Multiple Deprivation. Unadjusted and adjusted odds ratios (ORs) and 95% confidence intervals (CIs) calculated using logistic regression. Models adjusted for age only as a continuous variable. p-values by likelihood ratio test. Reference category for age group is aged 35 years and over