Food Standards Scotland
foodstandards.gov.scot

Food Standards Agency
food.gov.uk

E. coli O157 Super-Shedding in Cattle and Mitigation of Human Risk

November 2018

Contributing organisations

The Roslin Institute (RI): University of Edinburgh (UoE); Scotland's Rural College UK (SRUC) - SAC at the time of the proposal; ADAS UK Ltd: The Moredun Research Institute (MRI): The Scottish *E. coli* O157/STEC Reference Laboratory (SERL): Health Protection Scotland (HPS): Public Health England (PHE): The University of Glasgow (UoG): Animal Research Centre, United Stated Department of Agriculture (USDA-ARS)

Contents

Glossary 5

Executive Summary 7

Lay Summary 11

1. General Introduction 12

1.1 *Escherichia coli* O157 13

1.2. Colonisation of cattle by *E. coli* O157 14

1.3. Super-shedding from cattle 16

1.4. *E. coli* O157 subtypes and the link to super-shedding 17

1.5. Stx2 prophage association with super-shedding and human infection 18

1.6. Interventions to reduce excretion of *E. coli* O157 from cattle 19

1.7 Modelling interventions and the need for excretion and transmission data in relation to super-shedding strains. 20

1.8 Main Objectives 21

2. Objective 1: Farm *E. coli* O157 surveys and comparison of cattle and human isolates 22

2.1 SUMMARY 22

2.2 Two new structured surveys of farms in Scotland and England & Wales with kept cattle intended for the food chain (DO 2.1.1) 23

2.2.1. Introduction to the farm surveys 23

2.2.2 Methods: selection and sampling of farms 23

2.2.3 Results for the two new structured surveys 26

2.3 Comparative analyses of survey data (DO 2.1.1 cont.) 27

2.3.1 Level 1A: Comparison of current Scottish cattle E. coli O157 prevalence estimates to historical prevalence values 29

2.3.2. Level 1B: Comparison of current strain composition of Scottish cattle E. coli O157 with historical values (DO 2.1.4) 36

2.3.3. Level 2: Survey 3, Cattle only, Scotland versus England & Wales; A. prevalence; B. Strain composition (DO 2.1.4) 39

2.3.4. Level 3: Cattle versus human, Scotland versus England & Wales. A. prevalence; B. strain composition (DO 2.1.4) 43

2.4. Risk factors at farm level (final part of DO 2.1.4) 46

Risk factor analysis results 49

Interpretation of the outputs of the risk factor analysis 52

2.5 Sequence analysis of human and cattle *E. coli* O157 isolates (DO 2.1.2-3 & 2.2-6) 54

2.5.1. Human isolates for sequencing 54

2.5.2. Bovine isolates for Illumina platform short-read sequencing 54

2.5.3. Short-read sequencing and analysis methods 54

2.5.4. Results of phylogenomic studies based on short-read sequencing 55

2.5.5. Host and zoonotic prediction based on machine learning approaches 62

2.6. Whole genome sequencing (WGS) for routine diagnostics to enable SERL to transition to sequence-based diagnostics for *E. coli* O157 (DO 2.2.5-7) 62

2.7. Sequencing: insights from long-read sequencing of *E. coli* O157 isolates using the Pacific Biosciences platform (relates to DO 2.2.1-9) **62**
2.8. Conclusions from Objective 1 ..... **65**
3. Objective 2: Excretion dynamics and transmission frequencies of wild type *E. coli* O157 strains under controlled experimental conditions. **66**
3.1 SUMMARY **66**
3.2. Experimental design for excretion and transmission studies **66**
Characteristics of *E. coli* O157 strains used in experimental cattle studies **66**
Table 3.1. Details of *E. coli* O157 strains used for experimental calf studies. **67**
Animal study design **70**
Table 3.2. Summary of methods used to determine *E. coli* O157 shedding in cattle **72**
3.2. Experimental design for excretion and transmission studies **72**
Characteristics of *E. coli* O157 strains used in experimental cattle studies **72**
3.3 Results for Objective 2 ..... **72**
3.3.1. Excretion dynamics of PT21/28 and PT32 strains in orally challenged calves (DO 1.1.1-3 & 1.1.6-7). 72
3.3.2. Transmission of PT21/28 and PT32 strains between cattle (DO 1.2.1-4, 1.3.1/3) 74
3.3.3 Immune response to E. coli O157 in challenged calves (DO 1.1.4/5 & 1.3.2) 78
3.3.4. Contribution of Stx2a to survival within the ruminal gastro-intestinal tract (DO 1.3.3) 80
3.4. Conclusions from Objective 2 ..... **82**
4. Objective 3: To test a vaccine using a super-shedding strain and model the impact of on-farm interventions based on data generated within the programme **83**
4.1 SUMMARY **83**
4.2. Background to vaccine and modelling research **84**
4.3 Materials & Methods **85**
4.3.1. Data collection 85
4.3.2. Vaccination trial (DO 3.1.1) 85
4.3.3. Modelling approach (DO 3.2.1) 86
Removing the reservoir 89
4.4. Results **91**
4.4.1. Data collection 91
4.4.2. Vaccine trial data (DO 3.1.1-3) 92
4.4.3. Model parameters (DO 3.2.1) 93
4.4.4. Metapopulation simulations (DO 3.2.2-4) 97
4.5. Conclusions from Objective 3 ..... **99**
5. Outputs **101**
5.1. Recommendations **101**
5.1.1 Knowledge exchange to reduce the risk to humans 101
5.1.2 On farm control including vaccine field testing 101
5.1.3 Continued implementation of whole genome sequencing 101

5.2. Future Work 102
5.2.1. Mapping of human and animal isolates. 102
5.2.2 Further research on the role of Shiga toxins in ruminant and environmental reservoirs. 102
5.2.3 National surveys of sheep flocks and dairy herds. 102
5.2.4 Further research to understand vehicles of transmission including development of food testing methods. 103
5.2.5 Vaccine feedtrials 103
5.2.6 Risk factor analyses 103
5.2.7. Added value from projects 103
5.3. Knowledge Exchange 104
Output Publications (OP) 104
Scientific conferences: 105
National press 106
5.4. Personnel & Training 108
5.5. Added Value Projects 108
BECS added-value projects 108
6. Bibliography 110
APPENDICES 118
APPENDIX A: Sub-clustering and typing of *E. coli* O157: part 1 118
APPENDIX B: Deliverables as Direct Objectives (DO) of the programme 121
APPENDIX C: British *E. coli* O157 in cattle study manuscript 126
APPENDIX D: Additional data for Figures 2.4 and 2.6 127
APPENDIX E: WGS implementation at SERL and validation 130
APPENDIX F: Modelling approach used in Objective 3 133
APPENDIX G: BECS survey information provided to the participating farms 137

Glossary

Animal Health District. In the original SEERAD (1998-2000) survey Scotland was divided into six regions, based on Veterinary Animal Health Districts (AHDs): 1 = Islands; 2 = Highland; 3 = North East; 4 = Central; 5 = South East; 6 = South West. AHDs are no longer used within Scotland, however, they were retained within this study to maintain continuity.

Bacterial 'isolate' refers to the specific bacterium collected/isolated from a source, such as an animal or human. In a microbiological context, as we learn more about the properties of an isolate, it then can be referred to as a **strain**. Strain and isolate are often used interchangeably.

Bacteriophages (phages) are viruses that infect bacteria; they use the bacterium as a replication factory and kill the bacterium when they burst/escape from it. Some phages can have a lifecycle stage where they can integrate their genome into the bacterial genome (lysogenic) and hence their genetic code is copied along with the bacteria. Phages that encode Shiga-toxin (Stx) are usually lysogenic.

'**Cattle**' in this document refer to the bovids sampled or studied as part of this programme of work. For the prevalence surveys 'cattle' sampled were those destined for the food chain and the majority, but not all, were beef finishing cattle. Therefore some dairy cattle from mixed farms are included. By contrast, young male calves from dairy farms were generally used in the transmission and vaccine studies.

***Escherichia coli* O157** is a specific subtype of the *E. coli* species expressing the O157 LPS surface antigen. The majority of these strains contain genes for Shiga toxin (Stx) and can be referred to as Shiga toxigenic *E. coli* (STEC) O157.

Excretion vs. shedding. These terms are generally used interchangeably in this document to refer to the excretion/shedding of bacteria from the animal in faeces. We generally use 'shedding' more in an epidemiological context and 'excretion' in terms of faecal defecation.

Genome. *The core genome* refers to the genes present in nearly all members of a bacterial group (usually species). If you want to compare bacterial relatedness at the genetic level then you can analyse sequence variation in this core set of genes. *The accessory genome* refers to the genes that are variably present or absent in members of a bacterial group (usually species). The accessory genes are more likely to be present on prophages and plasmids that may vary between bacterial isolates.

Phage type (PT) in this report refers to sub-types of *E. coli* O157 determined by a laboratory method involving testing the survival or killing of *E. coli* O157 isolates when exposed to a panel of typing phages. PT21/28 is an important subtype in the UK associated with serious human infections.

Plasmids are self-replicating closed ‘circular’ DNA molecules that are separate from the main bacterial chromosome. Plasmids can vary in size from those that encode just a few genes to those encoding 100s of genes. They can be present in multiple copies in bacterial cells and usually be transferable between related bacteria.

Prophage is used to describe a phage genome when integrated into the bacterial genome in a lysogenic state. Mutations in such prophages can mean they become trapped in the bacterial genome and no longer can produce viable phages, these can be then termed cryptic prophages.

Shiga toxin (Stx) is the main virulence factor associated with the pathology during human disease and there are two main variants of the toxin Stx1 and Stx2. Each of these can be further sub-grouped: Stx 2a, b, c, etc. The toxin genes are encoded on integrated bacteriophage genomes within the *E. coli* genome and can be transferred between strains. Shiga Toxin producing *E. coli* are abbreviated as STEC.

Super-shedding is a term used to describe ‘high’ excretion levels of an organism such as a bacterium or virus from a host animal. In this context it refers to excretion of *E. coli* O157 in cattle faeces. The numerical cut-off for super-shedding has varied in different publications and so in the context of this report we have analysed data in relation to levels above 10^3 colony forming units (cfu) per gram of faeces. In most sections this has been further subdivided into ‘moderate’ (10^3 to 10^4) and ‘high’ (>10^4 cfu/g) shedding and these ranges were also tested in transmission models.

Type 3 secretion (T3S) is one of a number of different protein export system that can be encoded by bacteria. In the case of *E. coli* O157, the T3S system can export proteins into host cells that enable the bacteria to colonise in the host’s gastrointestinal tract.

Vaccine efficacy relates to how well a vaccine works in terms of the required criteria. In the case of *E. coli* O157 in cattle it refers to how well the vaccine performs in relation to both preventing colonisation of cattle and the extent to which it may reduce excretion of *E. coli* O157 if the animal is colonised.

Whole genome sequencing (WGS) for this study refers to the sequence of DNA bases of the bacterial genome of an isolate. This will be primarily be the sequence of DNA the constitutes the bacterial chromosome but will also include any plasmid DNA. Different methods for sequencing exist but in this report both ‘short-read’ Illumina and ‘long-read’ PacBio technologies were used.

Executive Summary

This research was funded to investigate *Escherichia coli* O157 high level excretion (super-shedding) from cattle and the threat this poses to human health. Specifically, the work was in four main areas:

(1) To investigate the prevalence of *E. coli* O157 in cattle destined for the food chain in separate surveys of farms in Scotland and England & Wales.
(2) To sequence cattle and human *E. coli* O157 isolates and determine their population structures and relationships with super-shedding and human disease.
(3) To determine the excretion dynamics and transmission frequencies between cattle of wild type *E. coli* O157 strains under controlled experimental conditions.
(4) To test an intervention using a super-shedding strain and model the impact of intervention based on data generated in this programme.

A significant part of the programme was based on surveys of the prevalence of *E. coli* O157 in faecal pats across 110 farms in Scotland and 160 in England & Wales which were completed between September 2014 and November 2015. This was the first study of *E. coli* O157 to be conducted contemporaneously across Great Britain, thus enabling comparison between Scotland and England & Wales. Although two previous surveys had been conducted in Scotland, no national survey has ever been conducted in England & Wales.

There was no statistically significant difference in either the herd-level prevalence estimates ($p=0.65$) or the pat-level prevalence estimates ($p=0.19$) for *E. coli* O157 in Scotland (herd level: 23.6%, 95% CI 16.6-32.5%; pat-level: 10.6%, 95% CI 6.7-16.3%) and in England & Wales (herd-level: 21.3%, 95% CI 15.6-28.3%; pat-level: 6.9(4.4-10.7%). The majority of *E. coli* O157 isolates (over 90%) were Shiga toxin positive, although the proportion of *E. coli* O157 isolates that were Shiga toxin negative was higher in England & Wales (17%) versus Scotland (<1%) ($p<0.001$). This might result in an overall lower prevalence of the more harmful STEC bacteria. A higher proportion of samples from Scotland were in the super-shedder category ($>10^3$ colony-forming units (cfu)/g of *E. coli* O157 in the faeces) compared to samples from England & Wales (2.4 fold more likely; $p<0.0003$). There was no difference between the surveys in the likelihood of a positive farm having at least one super-shedder sample. *E. coli* O157 continues to be common in British beef cattle, reaffirming public health policy that contact with cattle and their environments is a potential infection source.

Temporal analysis of the prevalence from all three Scottish national surveys showed that the farm-level prevalence was not significantly different across all 3 surveys (1998-2000; 2002-2004; 2014-2015). In Scotland approximately 21% (19%-24%) of farms were positive for *E. coli* O157 over this 17 year period. Previous farm status was not a predictor of current status. On farm pat-level prevalence, however, was significantly different across surveys with the second survey, IPRAVE 2002-2004,

being significantly lower than the other surveys. There were no spatial differences in prevalence but there were seasonal differences with higher values in the spring and winter. There has been little change across surveys in the strain composition as measured by Phage Type (PT). The PT21/28 *stx*2(a+c) strain remains dominant in Scotland, with *E. coli* O157 positive cattle on farms in Northern Scotland shedding entirely PT21/28. PT21/28, although still associated with higher level shedding (10^3-10^4), is no longer associated with super-shedding at levels >10^4 cfu/g faeces.

There were spatial differences in strain composition (as measured by PT) across Scotland, England & Wales. In Scotland cattle are primarily shedding PT21/28, especially on northern farms, where 80-100% of the positive cattle were shedding this subtype. By contrast, in England & Wales PT21/28 was only found on 3 farms located in the North East, Wales and West Midlands. The diversity of O157 PTs was much higher in cattle in England & Wales compared with Scotland. The number of UK PTs recorded, varied from regions with >4 PTs represented (e.g. West Midlands) to regions where only one PT was represented (e.g. Yorkshire).

Data from health protection agencies in Scotland (HPS/SERL) and England & Wales (PHE) indicate that although the trend is downwards, there has been no significant decrease in the rates of reported human clinical cases in Great Britain since 1998 with the rates in Scotland consistently 3-fold higher than England & Wales. Our comparative analysis of *E. coli* O157 isolates from both cattle and humans indicates that related cattle isolates can be identified for the majority of human isolates, supporting the generally accepted concept that many of the human *E. coli* O157 infections in the UK originate from a cattle source. However: (1) a subset of human infections do not match with national cattle isolates and may be explained by travel and non-UK food consumption and/or an alternative un-sampled reservoir; (2) cattle specific clusters are also observed which contain strains not currently associated with human infections. Our data affirms that while cattle are a major reservoir for Shiga toxin positive *E. coli* O157 strains that can infect humans, the vehicles and routes for human infection are still only partially understood.

Whole genome sequencing (WGS) was used throughout the project to aid with epidemiological analysis of *E. coli* O157 isolates. Under the programme, we helped install a sequence analysis pipeline at the Scottish *E. coli* Reference Laboratory (SERL), developed at PHE, to facilitate *E. coli* O157 diagnostics and outbreak investigation. Sequence data is powerful as it allows precise relationships between isolates to be determined and therefore outbreaks to be accurately defined and sources potentially attributed during an outbreak. For example, whether a human infection is likely to have arisen from the local farmed animal population or by a strain present in imported food or as a consequence of travel abroad. For *E. coli* O157 subtype PT21/28, we demonstrate that the majority of human isolates can be traced back genetically to recent ancestors present in the British cattle population.

Our WGS analysis of both historical and recent isolates from cattle and humans has helped understand the relatively recent appearance of this potentially lethal zoonotic disease and places our isolates into an international context. We provide evidence that the emergence of serious human *E. coli* O157 infections coincides with the introduction of a specific sub-type of Shiga toxin (Stx2a) into our cattle reservoir. While there are multiple types of *E. coli* O157 with this toxin in England & Wales, in Scotland there is a single predominant strain type (PT21/28) producing this toxin and this has been present for the last decade. This PT is associated with serious human infections and potentially explains why Scotland has higher rates of human infections with *E. coli* O157 than England & Wales.

We have used the human and cattle isolate sequence data with machine-learning approaches to investigate both 'source attribution' of *E. coli* isolates and to predict which *E. coli* O157 isolates pose the greatest threat to human health. Our work indicates that such methods can accurately predict the host or environmental niche that an *E. coli* comes from. This has value for investigating the source of contaminating bacteria in water-courses, food and human infections. The capacity to recognise the *E. coli* O157 isolates that pose the greatest risk to human health could allow more targeted interventions in livestock reservoirs, for example the use of a vaccine on farms with such high risk strains.

Under controlled experimental conditions, we have demonstrated that a PT21/28 *E. coli* O157 isolate is excreted from cattle at higher levels than an alternative subtype, PT32. These higher levels of shedding of the organism may mean it is more likely to colonise other animals and contaminate food/water for human consumption. We have demonstrated that the presence of Stx2a plays an important role in establishing high level excretion in nearby exposed animals and so this toxin is important for maintaining specific *E. coli* O157 strains in the cattle reservoir through enhanced transmission. The reason behind the advantage conferred by Stx2a in the ruminant reservoir still remains to be determined. Our preliminary experiments did not support a published hypothesis that Shiga toxin a (Stx2a) improved transmission and colonisation of cattle due to killing of grazing protozoa in the rumen. A PhD study under this programme also did not find clear evidence that the toxin provided an advantage through immune response manipulation. On-going research will focus on its role in aiding *E. coli* O157 attachment in the bovine gastrointestinal tract by altering receptor distribution and inhibiting epithelial cell turnover.

Long-read sequencing has provided valuable insights into the evolution of *E. coli* O157, in particular to show the importance of changes outside the core genome, for example in prophage elements and plasmids. Such sequencing shows the changes that can occur in closely related *E. coli* O157 isolates over very short time scales. This was exemplified by comparison of related isolates from two outbreaks that occurred within months at the same restaurant. Plasmid acquisition and large-scale genome re-arrangements between the isolates altered factors that could impact on

the likelihood and severity of human infection. Our long-read sequencing data indicates that large scale genome re-arrangements are common in *E. coli* O157 PT21/28 and we speculate that this genome plasticity may confer an adaptive advantage to isolates and therefore contribute to its particular success as a zoonotic pathogen.

Finally, a vaccine formulation we have developed over the last eight years based on a combination of *E. coli* O157 antigens involved in colonisation, was shown to reduce significantly excretion levels of a *stx*2a+ *E. coli* O157 isolate from cattle as well as limit its transmission to other in-contact calves. We have used existing models developed to capture the transmission of *E. coli* O157 amongst Scottish cattle and extended these to incorporate a deeper understanding of the super shedding dynamics derived from the experimental trials conducted in this study. Data from these trials was also used to provide vaccine efficacies for these models. Our models show that annual vaccination and vaccination at the point of movement should both be effective, but that annual vaccination should have greater and more rapid impact. Our results show that starting with a conservative herd-level prevalence of 28% and using models parameterised for the most transmissible strain (the repaired PT21/28 strain) vaccinating annually is predicted to reduce herd-level prevalence to below 1% (95% CI: 23 – 0 %) within 3 years.

Lay Summary

Cattle are a reservoir of a specific type of *E. coli* bacteria known as *E. coli* O157. These bacteria can produce toxins (Shiga toxins) that can cause life-threatening human infections. Previous work has shown that there are different subtypes of *E. coli* O157 and that the particular subtypes excreted from the cattle in the highest amounts are those which are directly associated with the most severe human infections. The work in this programme had four main objectives:

(1) To examine how common *E. coli* O157 is across cattle farms in both Scotland and England & Wales.

(2) To then compare (by genome sequencing) the *E. coli* O157 subtypes isolated from cattle with those isolated from human infections to understand if specific subtypes in animals pose more of a threat to human health.

(3) To test if high level excretion (known as 'super-shedding') is associated with the bacteria being able to produce a specific type of Shiga toxin, and

(4) To test a vaccine developed to limit *E. coli* O157 excretion from cattle and determine if its use can be a way to prevent transmission of the bacteria between cattle and humans.

Our research has established that levels of *E. coli* O157 in cattle have remained relatively constant in Scotland over the last decade and are equivalent to those in England & Wales. Approximately 20% of farms and 10% of animals were positive for *E. coli* O157 based on faecal pat sampling.

The diversity of *E. coli* subtypes in cattle was, however, much greater in England & Wales compared to Scotland which had a high level of a certain subtype associated with super-shedding and severe human infection. Local exposure to this subtype may explain the higher human incidence in Scotland compared to England & Wales.

The accuracy of whole genome sequencing helps define which specific subtypes are associated with an outbreak; it can help trace the origin of an outbreak and can be used to make predictions about the threat to human health posed by specific isolates.

The research established that Shiga toxin subtype 2a is important for transmission of *E. coli* O157 between cattle as it enabled bystander animals in a group to become colonised following introduction of an animal excreting *E. coli* O157. The trialled vaccine limited both excretion from cattle as well as transmission to cattle and our modelling indicates that such a vaccine should be of significant benefit in terms of reducing human exposure and infection from *E. coli* O157.

1. General Introduction

Escherichia coli is a common bacterium that can be isolated from the gastrointestinal tracts of humans and many mammals, birds and reptiles. It is a facultative anaerobe which means it can grow in the presence or absence of oxygen and is able grow on a wide range of carbon and nitrogen sources. It is simple to isolate on specific agar plates and this quality has enabled its use as an indicator of faecal contamination, for example in water.

It is a very diverse species as it has an 'open genome'; this means it is constantly losing and acquiring genes. The majority of *E. coli* strains are commensal in nature i.e. they are not harmful to a healthy individual. However, a subset of *E. coli* strains are more virulent and can be considered overt pathogens. *E. coli* can cause infections associated with the gastrointestinal tract or 'extra-intestinal'; such as in the urinary tract or bloodstream. Some *E. coli* isolates, especially those present in cattle, can be lysogenised by phages that encode different types of Shiga toxin (Stx) and therefore produce these toxins. A subset of the these 'Shiga toxin producing *E. coli*' (STEC) also have genes encoding factors that allow infection of the human intestinal tract, especially a type 3 secretion system (T3SS). Strains that encode Stx and a T3SS and have caused disease in humans can be defined as enterohaemorrhagic *E. coli* (EHEC). EHEC strains can have different surface properties which are detected by serological tests and are then referred to as serotypes. *E. coli* O157 is a common EHEC serotype, especially in the United Kingdom (UK).

E. coli O157 first appeared as a significant zoonotic threat in 1982 associated with cases of unusual gastrointestinal disease in Oregon and Michigan. These infections were linked to a beefburger restaurant chain. Since this time there have been a number of important incidents including the infamous 'Jack-in-the-box' outbreak that had long-standing repercussions for food safety in the United States of America (USA). In the UK, it was the 1984 outbreak stemming from a butcher's shop in Lanark that announced its lethal presence, with ~20 associated deaths, and after the published enquiry into the outbreak there was significant investment in research to understand both the prevalence of *E. coli* O157 in cattle in Scotland and fundamental research into the biology of the organism. Despite a now sophisticated understanding of this pathogen, the levels of *E. coli* O157-associated human disease in the UK have remained relatively constant over the subsequent 25-year period.

E. coli O157 is a zoonosis that originates in ruminants, in particular cattle, but causes no obvious disease in the ruminant host. As such where the responsibility lies to prevent or limit human infections is still a source of contention. This research was commissioned by Food Standards Scotland (FSS) and the Food Standards Agency (FSA) in response to recommendations from an *E. coli* O157 outbreak in Wales[1] and following consultation with stakeholders at a specific workshop held in 2011[2]. It was evident from the workshop that serious human infections normally originated from the sub-set of *E. coli* O157 isolates that were excreted from cattle at high levels

(super-shedders) and therefore a greater understanding of this link was required. In response, proposals were invited for research that would:-

> '*Improve the understanding of E. coli O157 excretion by cattle and the super-shedding phenomenon, and how this pathogen can be controlled on-farm. The FSA is seeking applied, useable outputs from this package of work which can be used to inform future on-farm control strategies for reducing the public health risk associated with this pathogen in the UK'*.

1.1 *Escherichia coli* O157

Human infection with enterohaemorrhagic *E. coli* O157 can be fatal or result in life-long morbidity including brain and kidney damage[3]. Based on our current understanding, ruminants are the main host reservoir for *E. coli* O157 strains and in the UK, *E. coli* O157 is the main serogroup associated with human disease. The serious pathology associated with *E. coli* O157 infections in humans is a consequence of the activity of Shiga toxins (Stx) which are released in the gastrointestinal tract by the bacteria. Stx can traffic across the epithelial barrier in the gut and can kill cells lining human blood vessels. The response to this toxin-mediated damage can further contribute to this life-threatening infection by causing damage to red blood cells, a condition known as haemolytic uraemic syndrome (HUS). *E. coli* O157 is the leading cause of acute paediatric renal failure and HUS in the UK, and patients may require life-long dialysis or a kidney transplant as a consequence[1].

In Scotland, the first reported cases of human *E. coli* O157 infection were identified in 1984. Currently, Health Protection Scotland (HPS) conducts active, population-based enhanced surveillance in close collaboration with the Scottish *E. coli* O157/STEC Reference laboratory (SERL). Over the 14-year period preceding this research application (1998-2012), HPS reported a mean of 224 (205-243) culture-positive cases of *E. coli* O157 at an annual rate of 4.37 (4.02-4.71) cases per 100,000 of the population. The figures for the UK from 2008-2017 are shown in Fig. 1.1. Rates of human *E. coli* O157 infection in Scotland are generally higher than in most other UK, European and North American countries. Of the ~200 cases per year in Scotland, over 40% required hospitalisation and almost 10% developed severe renal complications. The severe effects of the infection are more likely to occur in young children.

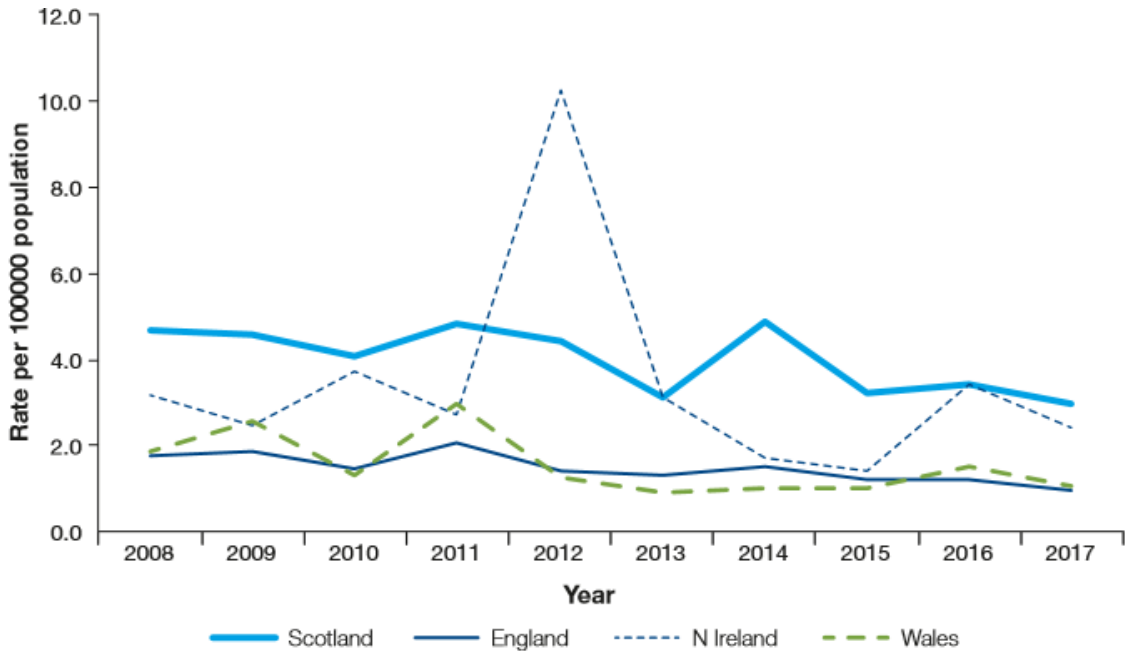


Figure 1.1. *E. coli* O157 rates of per 100,000 population in the UK 2008-2017*. Culture positive cases from the *Surveillance Report STEC in Scotland 2017: Enhanced Surveillance and Reference Laboratory Data*. hps.scot.nhs.uk/resourcedocument.aspx?id=6745

There are numerous subtyping schemes for *E. coli* O157 which are valuable for epidemiological studies and outbreak investigation. There are three main lineages of *E. coli* O157 (I, I/II & II) but also subdivision by clades (1-9)[4,5]. We have provided APPENDIX A which shows *E. coli* O157 phylogeny based on SNP variation and the association with clustering by lineage and clade. It also highlights some of the main clusters by specific countries that have significant human infection issues with *E. coli* O157. In the UK, a common way to sub-type *E. coli* O157 strains for relatedness has been phage typing, for which the susceptibility of a specific isolate is determined against a bank of 16 lytic typing phages. There is a good relationship between phage type (PT) and clustering from WGS data (APPENDIX A).

1.2. Colonisation of cattle by *E. coli* O157

In 2003, our research grouping demonstrated that *E. coli* O157:H7 primarily colonises the terminal rectum of cattle[6] (Fig. 1.2). While other sites in the gastrointestinal tract may also be colonised, by far the highest levels of replicating organism were recovered from this specific site. This tropism has been confirmed for multiple strains and from naturally and experimentally-colonised animals[7–9]. As a consequence, when the infected animal defecates, the faeces is coated with mucus containing high levels of *E. coli* O157. We consider that 'super-shedding', excretion of high levels of the bacterium (>10^3 per gram of faeces) requires rectal colonisation to allow high number of bacteria to replicate and then be excreted. We can reproduce this colonisation and super-shedding in calves following oral inoculation of

E. coli O157. Subsequent studies have identified multiple bacterial factors important for colonisation, including the type III secretion system (T3S) and Shiga toxin encoding bacteriophages[10–12]. The bacterial T3S system injects a cocktail of about 40 different effector proteins into the host epithelial cell and these facilitate binding to the cell and manipulation of innate signalling responses[13–15]. Expression of T3S is variable between strains and we have also demonstrated that the repertoire of Stx-encoding bacteriophages integrated into the *E. coli* O157 strain impact on the regulation of type III secretion[11,16,17].

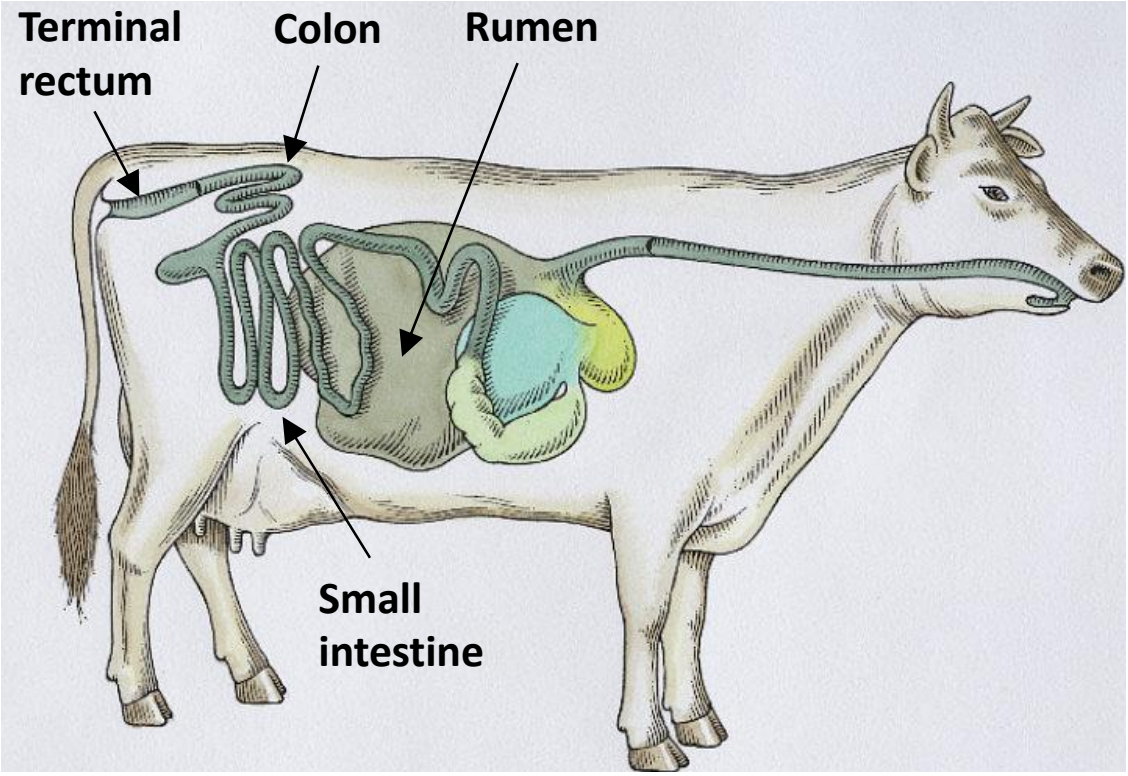


Figure 1.2 Diagram illustrating the intestinal tract of a cow. *E. coli* O157 predominately colonises the terminal rectum of the animal.

We have developed a model to explain how bacterial colonisation and replication rates interact with the host's response leading to the patterns of excretion we measure following experimental infection in cattle[18]. Stx has been shown to repress innate responses and inhibit the proliferative capacity of local B and T cells, thereby interfering with adaptive host responses[19–21]. Stx can also aid *E. coli* O157 colonisation by promoting expression of a host cell receptor for the bacterial adhesin intimin[22,23] and a genome-wide screen identified *stx* as a factor significant for colonisation of *E. coli* O157 in young calves[24]. As a consequence, variation in the types and level of Stx expressed by *E. coli* O157 are likely to be linked to levels of *E. coli* O157 excreted from cattle, including super-shedding. It has also been proposed that Stx kills grazing protozoa that predate on *E. coli* in the rumen[25,26]. In summary, Stx function and regulation have evolved in the bovine host to promote bacterial colonisation but when humans are infected with *E. coli* O157, Stx toxin expression can be life-threatening.

1.3. Super-shedding from cattle

Following investigations of outbreaks, it is widely considered that ruminants, in particular cattle, are the primary source of the *E. coli* O157 strains that infect humans. Infections used to be mainly associated with consumption of undercooked or raw meat but this has shifted to include exposure through consumption of contaminated foods and drinks, such as salads, unpasteurised cheese, and fruit juices, as well as direct contact with animals and their environment. A recent WHO report has detailed food-based attribution, hazard and monitoring for STEC including *E. coli* O157[27]. While *E. coli* O157 has been detected in a wide spectrum of other animal species (sheep, goat, deer, moose, swine, horse, dog, cat, pigeon, chicken, turkey, gull), sometimes even with considerable prevalence[28,29], it is evident that ruminants harbour multiple *E. coli* strains that encode Stx. Specific trace-back studies have isolated the same strains in cattle that have caused serious disease in humans.

In a 2008 publication we proposed a specific mechanism for the link between human infection and livestock carriage of *E. coli* O157 that involved a subset of shedding animals known as super-shedders[6,30]. Super-shedders are individuals who, for a defined period (usually only a matter of days), yield more infectious organisms (here *E. coli* O157) than typical individuals of the same host species[30]. The level defined in that publication for super-shedding was 1×10^4 cfu/g faeces. While most subsequent work has made use of 10^4 as a cut-off, a level of 10^3 for super-shedding has also been proposed in a recent analysis[31]. In the current study we have carried out analyses with both thresholds including a 'moderate' category from 10^3-10^4. Shedding high concentrations of *E. coli* O157 has been proposed as a major contributor to cattle-to-cattle transmission and possibly cattle-to-human transmission.

In an international context there have been a number of studies that have analysed risk factors associated with super-shedding. There is some evidence that the gastrointestinal tract microbial community may be different depending on super-shedding status although so far these experiments have been very limited in animal numbers analysed[32,33]. The animal's diet can impact on shedding of *E. coli* O157 which will also inevitably alter the faecal population structure, specific factors such as distillers' grains have been shown to increase excretion levels of *E. coli* O157[34,35]. Cattle factors have been investigated by looking at gene expression with preliminary data showing some association of super-shedding with T-cell responses and cholesterol metabolism[36,37].

Research centred in Scotland has been a focal point for understanding the biology and epidemiology of super-shedding for *E. coli* O157[6,38–40]. The central tenet of our proposed research is that not all *E. coli* O157 strains are equally likely to cause super-shedder infections and this then relates to the isolates more likely to be associated with disease and potentially disease severity in humans.

1.4. *E. coli* O157 subtypes and the link to super-shedding.

Two surveys of Scottish store and finishing beef cattle were conducted between 1998 and 2004. The SEERAD survey (funded by the Scottish Executive Environment and Rural Affairs Department – SEERAD 1998-2000) and IPRAVE survey (funded by a Wellcome Foundation International Partnership Research Award in Veterinary Epidemiology 2002-2004) represent the only reported systematic national surveys of bovine *E. coli* O157 shedding and present a valuable opportunity to examine changes in patterns of shedding and strain characteristics. 14,849 faecal pats across 952 farms were sampled in the SEERAD survey and 12,963 pats across 481 farms in the IPRAVE survey. A total of 1,296 *E. coli* O157 strains were isolated from the SEERAD survey (n=207 farms) and 513 strains in the IPRAVE survey (n=91 farms). In both cattle and humans, the predominant phage type (PT) between 1998 and 2004 in Scotland was PT21/28, which comprised over 50% of the positive cattle isolates and reported human cases respectively; thus providing evidence of a link between bovine excretion levels and human infection. In contrast, the proportion of another subtype, PT32, was represented by relatively few (<5%) human cases in the UK despite comprising over 10% of cattle isolates (Fig. 1.3A).

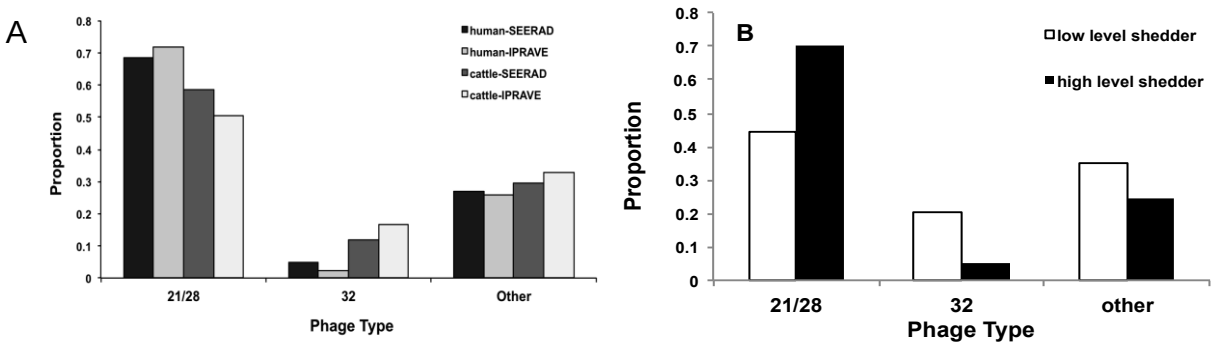


Figure. 1.3. **Phage type and relationship with host and super-shedding.** (A) Proportion of PT21/28, PT32 and 'other' PTs in cattle isolates and in culture positive, non-travel related human *E. coli* O157 cases with known PT results reported to HPS, over the time periods equivalent to the SEERAD (March 1998 – May 2000) and IPRAVE (February 2002 – February 2004) surveys[40]. (B) PT and distribution into high shedders (>10^4) and low shedders (<10^4)[4]. PT21/28 is significantly associated with high-level (>10^4) shedding (chi-square=6.54; p=0.015; odds ratio=2.90 (1.18-7.83)).

From the IPRAVE survey, 440 strains had associated count data from faecal pats allowing identification of high-level shedders (super-shedders) and low-level shedders (non super-shedders). Results revealed that the proportion of high-level shedders that were PT21/28 was higher than expected by chance (p=0.015). The odds of shedding PT21/28 were over twice as high for high-level shedders than low-level shedders. The proportion of PT32, however, was lower than one might expect from chance alone among high-level shedders compared with low-level shedders

(Fig. 1.3B). This supports the contention that phage type PT21/28 is more transmissible from cattle to both other cattle and humans compared to PT32.

1.5. Stx2 prophage association with super-shedding and human infection

A preliminary genomic comparison of the Scottish PT21/28 and the PT32 strains showed that the major difference between them is the presence of an Stx-encoding bacteriophage inserted at tRNA-*arg* encoding the *stx*2a allele and an 8.8kb additional phage region within the tRNA-*leu* insertion site of strain 9000 (PT21/28). This provisional genome analysis is in line with our published work[11,41], that demonstrates that the majority of PT21/28 strains contain both Stx2a and Stx2c encoding prophages whereas the majority of PT32 strains in cattle encode just Stx2c. While the situation appears similar in humans for the PT21/28 strains, there is a significant shift in the distribution of Stx types in the PT32 strains associated with human infection (Fig. 1.4B), with the majority containing a Stx2a-encoding prophage. Based on this finding, Stx phage type was analysed retrospectively by PCR across 143 IPRAVE cattle strains and the association with shedding level determined. Strains associated with higher pat levels are significantly more likely to contain *stx*2a[41] (Fig. 1.4A). This is in agreement with multiple international reports showing that human infection is more likely to be associated with particular *E. coli* O157 strain lineages and *stx* type[5,42–44].

As preliminary data for our application to this funding call, two strains were selected for challenge studies in calves. The first was PT21/28 strain 9000 isolated from a bovine faecal pat with a high level of *E. coli* O157 (686,400 cfu/g) and this strain contains both Stx2a and Stx2c prophages. The second was PT32 strain 10671 isolated from a bovine faecal pat with <100 cfu of *E. coli* O157. This strain contains only a Stx2c prophage. In the pilot study, four calves were orally dosed with 10^9 cfu of each strain in separate rooms that also contained two uninfected sentinel animals to provide an indication of strain transmission under these conditions. Excretion was monitored in faeces over a four-week period, with the four sentinel animals also becoming positive within the first three days, although for one PT32 animal this was very transient. Statistical analysis (of the area under the bacterial shedding curve, AUC, which represents total shedding of *E. coli* O157 over the duration of the study period) showed that the PT21/28 strain was excreted at significantly higher levels over the duration of the experiment and this was true for both the main orally inoculated group and the sentinel animals (Fig. 1.4C). Therefore, it is evident that the two strains appear representative of the epidemiological data with higher shedding exhibited by the PT21/28 isolate. This work also demonstrates that we can use these strains under controlled colonisation conditions and get effective transfer to sentinel animals from which to obtain excretion curves in line with natural colonisation as opposed to those following high dose oral administration.

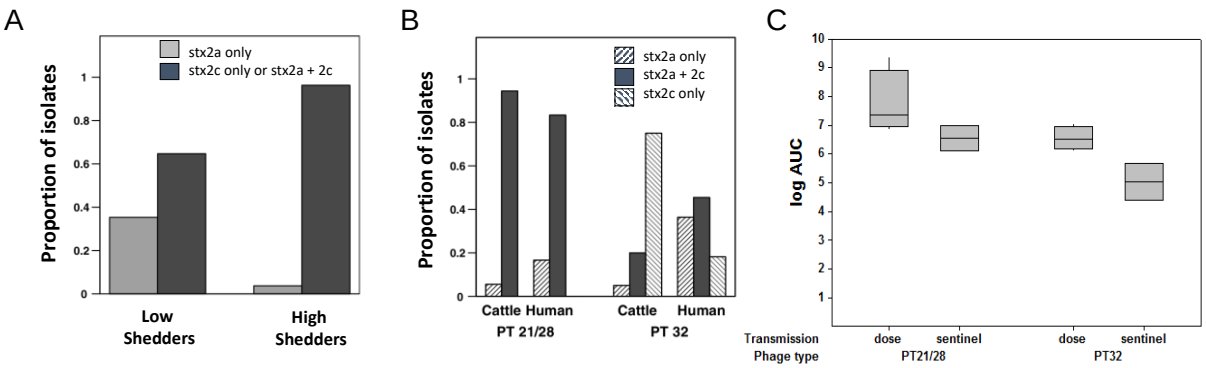


Figure 1.4. Relationship between shedding and phage type in cattle. (A) The majority of high shedders (total number of strains analysed n=143) were *stx2a*+ (B) PT21/28 *stx2a*+ was the most prevalent O157:H7 phenotype in cattle and humans and there were higher levels of the *stx2a*+ PT32 strains in humans compared to cattle (C) Following oral challenge of calves (n=4 per strain), total shedding of bacteria, as determined by AUC, was significantly higher for PT21/28 (*stx2a, stx2c*) strain 9000 compared to PT32 (*stx2c* only) strain 10671 (p=0.019). *E. coli* O157 naïve sentinel calves (n=2 per strain) co-housed with orally challenged calves also showed a difference in shedding level between the two phage types, with shedding higher in sentinels calves exposed to calves shedding PT21/28 strain 9000.

1.6. Interventions to reduce excretion of *E. coli* O157 from cattle

A variety of different animal-targeted interventions have been investigated which include drinking water treatments, dietary manipulations, probiotics or other feed additives, improved farm biosecurity, use of bacteriophages and vaccines[45–48]. A systematic review[47] of these pre-slaughter interventions found only probiotics and vaccines to be effective at reducing shedding of *E. coli* O157 in ruminants. This was in agreement with a 2011 FSA commissioned report on the feasibility of introducing methods in the UK for reducing shedding of *E. coli* O157 in cattle. The report concluded that probiotics and vaccines (neither of which are currently commercially available in any country) were the most promising interventions and would have similar costs to implement in UK cattle[49]. However, probiotics need to be administered on a daily basis[50] and are therefore difficult to implement in UK cattle populations, which are generally extensively reared (i.e. not housed) during summer and autumn periods where *E. coli* O157 shedding is highest. In contrast vaccines typically require only two to three doses and are therefore more easy to use in non-housed cattle. We therefore chose to focus on vaccine-based interventions within this project.

A systematic review of cattle *E. coli* O157 vaccines identified those that target bacterial adherence to the gut wall and bacterial iron sequestration systems as the

most efficacious to date[46]. Only two vaccines that claim to limit *E. coli* O157 shedding from cattle have obtained provisional licences. One of these was developed by Bioniche Animal Health and is comprised of concentrated *E. coli* O157 culture supernatant containing a cocktail of proteins, including those secreted by the bacterial T3S. We have worked extensively on the regulation and function of the *E. coli* O157 T3S system and have shown it is essential for colonisation of cattle[10,24]. Taken together, this led us to evaluate subunit vaccines based on T3S components which have been shown to be effective under experimental conditions[51–54]. The second vaccine was developed by Epitopix and recently licenced by Pfizer Animal Health. This relies on a 'siderophore receptors and porin' (SRP®) preparation comprised of a mixture of membrane proteins extracted from bacteria cultured under iron-limiting conditions. The Epitopix vaccine has been tested under both experimental and field challenge conditions, including larger scale trials in North American feedlot cattle[55,56] and there is considerable variability in how this vaccine has performed under field conditions, suggesting that further work needs to be carried out to define a successful vaccine formulation and how vaccines can be best used in practice[55,56].

1.7 Modelling interventions and the need for excretion and transmission data in relation to super-shedding strains.

While experimental studies of the currently available *E. coli* O157 vaccines have largely focused on the ability of the vaccines to reduce bacterial shedding from experimentally-infected cattle, field evaluations of these vaccines have generally examined their capacity to reduce prevalence of *E. coli* O157 at the herd infection level. However, our previous work provides evidence that shedding intensity in cattle has two important roles. Firstly, that high shedding increases transmission between cattle. By fitting mathematical models to field data, we have shown that high intensity shedding is a major driver of cattle to cattle transmission[38,41] and cattle shedding the super-shedder strain, PT21/28, have over double the transmission rate of cattle shedding a PT32 strain[40]. Secondly, we have examined high intensity shedding as a source of risk to humans and by comparing the frequencies of super-shedder and non super-shedder strains in the cattle reservoir and in humans, we were able to show that human risk was dominated by shedding above a threshold of around 1,300 cfu/g[41]. Therefore, in order to predict how alternative vaccination strategies would reduce prevalence and total shedding at a herd level and how effective they might be at reducing human risk requires data on shedding intensity and transmission rates in naturally infected, vaccinated and unvaccinated animals are required. These parameters have been derived as part of this project using calf sentinel experiments.

These parameters have been used in meta-population models of the national cattle population that are currently being developed under the Scottish Government funded SPASE project (Scottish Partnership for Animal Science Excellence:

https://www.gov.scot/Publications/2017/09/9471/17). The models capture stochastic transmission dynamics at the herd level coupled with infection from environmental sources and via movements of infected animals. The data generated in these projects will enable us to parameterise the time course of infection and transmissibility in individual animals. Using these parameterised models we will be able to examine how alternative vaccination strategies – differing for example in proportion of uptake by farmers – would reduce the national prevalence of infection in cattle and reduce the number of human infections.

1.8 Main Objectives

1. Surveys of the prevalence of *E. coli* O157 across cattle farms in England & Wales and Scotland. *E. coli* O157 will be isolated from cattle faecal pats, typed and a subset will be sequenced. These will be compared with *E. coli* O157 isolates from human infections in the UK over the same time period to determine the subtypes in cattle that are a threat to human health.

2. To determine *E. coli* O157 factors that contribute to super-shedding from cattle as well as transmission between animals. This will be done by experimental challenge of calves with strains that have different properties. A specific focus will be the role of Shiga toxin subtype 2a (Stx2a).

3. To test a vaccine formulation using a super-shedding strain and model the impact of on-farm interventions such as vaccines based on data generated within the programme.

2. Objective 1: Farm *E. coli* O157 surveys and comparison of cattle and human isolates

2.1 SUMMARY

This objective included two new structured field surveys; one in Scotland and one in England & Wales which enabled comparison with previous cattle surveys in Scotland. The isolation and characterisation of *E. coli* O157 strains from cattle also allowed comparison with human *E. coli* O157 isolates from the same time period.

- *E. coli* O157 was found in faecal pat samples collected from cattle closest to finishing on approximately one in five of the sampled farms. This confirms the importance of cattle and their environment as reservoirs for *E. coli* O157.
- Approximately 24% of Scottish farms were positive for *E. coli* O157. There was no significant difference in the proportion of farms that were positive for *E. coli* O157 across the three national cross-sectional surveys carried out in Scotland (1998-2015).
- Approximately 21% of the England & Wales farms were positive for *E. coli* O157. There were no significant differences in either the herd or the pat-level prevalences of *E. coli* O157 between Scotland and England & Wales, but the strain types and the proportion of *stx* negative O157s were very different.
- In Scotland, PT21/28 was still the dominant PT in cattle, and 65% of positive farms (n=17/26) had cattle shedding PT21/28. It is still associated with higher shedding levels (10^3-10^4), although not with shedding at >10^4 as it was during the two previous national cross-sectional surveys, SEERAD and IPRAVE. In some areas of North East Scotland, all positive *E. coli* O157 isolates from cattle were PT21/28. By contrast, PT21/28 was not the dominant PT in cattle in England & Wales, as only 9% of positive farms (n=3/34) had cattle shedding PT21/28. There is no dominant PT in cattle in England & Wales and there is no one singular strain sub-type that is associated with high shedding
- In Scotland, the rate of reported human clinical cases continues to be much higher (by approximately 3-fold) than in England & Wales. Both PT21/28 and *stx*2a have been associated with higher severity of illness, which may explain the differences observed in human cases. Assuming that cattle are the main reservoir, cattle in England & Wales have less PT21/28 and *stx*2a than cattle in Scotland.
- Risk factor analysis: the surveys were not formally designed to investigate risk factors and can only suggest factors associated with: (1) farm *E. coli* O157 status; (2) the proportion of samples positive within positive groups or (3) the presence of a super-shedder. In Scotland, season and some management/demographic characteristics were associated with farm O157 status, including moving breeding females onto the farm and purchase of livestock other than cattle in the year prior to sampling. The total number of cattle aged 12-30 months on the farm was significant in Scotland as was larger herd size in England & Wales. In both surveys, being housed was positively associated with *E. coli* O157 status but potentially confounded by season. Groups with at least one super-shedder were also more likely to have a larger proportion of samples testing positive than groups without a super-shedder.
- We helped implement routine WGS for *E. coli* O157 at the Scottish *E. coli* O157/STEC Reference Laboratory (SERL). In terms of public health, sequencing isolates from both farm surveys and human infections provides a major advance in understanding what strains of *E. coli* O157 are likely to be from a domestic reservoir and supports outbreak investigations. Machine learning based on this data can predict cattle *E. coli* O157 isolates more likely to be a threat to human health.

As this objective contained multiple sub-objectives that have required separate reporting it has been divided up into 4 subsections (2.2-2.6) each with an introduction, methods and results.

2.2 Two new structured surveys of farms in Scotland and England & Wales with kept cattle intended for the food chain (DO 2.1.1)

APPENDIX B contains the table of deliverables/direct objectives (DO) for the project

The two new structured surveys represent a major component of the programme and they allow comparison of data with previous survey data from Scotland. The methods and primary analysis of the two new surveys (otherwise known as the British *E. coli* O157 in cattle Study, or BECS) have been published[57] (Output1-OP1: Henry *et al* 2017). This report will first provide details about the two new surveys including methods and results, and this is then followed by a comparative analysis of the survey data from the different time frames (2.3).

2.2.1. Introduction to the farm surveys

From September 2014 to November 2015, two structured field surveys in cattle were conducted; one in Scotland and one in England & Wales. The primary objective of these surveys was to estimate how many herds of cattle were likely to have at least one animal shedding *E. coli* O157 in their faeces during that time period. Another was to see how many samples (pats) were positive for *E. coli* O157 out of all the samples collected (giving a pat level prevalence). These values have previously been estimated for Scottish cattle herds in two studies; one in 1998-2000 (SEERAD)[58] and one in 2002-2004 (IPRAVE)[40]. The numbers of *E. coli* O157 bacteria found in those samples that were positive were counted and one confirmed *E. coli* O157 isolate per positive sample was archived for further use.

2.2.2 Methods: selection and sampling of farms

Scotland survey: A number of Scottish farms had participated in two earlier *E. coli* O157 surveys – one carried out between 1998 and 2000 and another between 2002 and 2004. Samples from at least 110 farms were required for this new survey. These were randomly recruited from farms that were believed to have participated in both the previous surveys. However, it was later (during the analysis in 2.3) that we discovered that eight of these had apparently not been included in the analysis of both the previous surveys. These 110 farms were visited between September 2014 and September 2015. Initial postal contact was made with all the farmers that had participated in both of the original Scottish surveys, who were known to still be in business as registered cattle keepers and who were recorded as having at least one bovine animal in at least one of four Cattle Tracing System categories (CTS 303, 307, 310, 311) on their farm on June 1st 2013. These categories contain cattle aged between one and two years and cattle over two years without offspring. These

categories were the most relevant because the target animals were groups of cattle that would end up in the food chain, so farms with cattle in these categories would be suitable for the survey. The initial contact letter provided information about the survey and gave farmers the chance to opt out of taking part. Those farmers who had not opted out were then phoned, in a randomised order, to arrange a visit to their farm at a convenient time.

On each farm, samples of faeces were collected from fresh pats that had been passed by the group (or groups) of cattle that were closest to being sent to slaughter. The number of samples taken depended on the number of cattle in the group. This number followed pre-planned protocols[57] that were developed in the earlier Scottish surveys. The protocols were designed to ensure a 90% probability of detecting *E. coli* O157 in the group of cattle, if at least one animal were shedding the bacteria in its faeces. The faeces samples were put into pots, then packaged and sent by courier to the microbiology laboratory in Inverness.

The SRUC project team was responsible for farmer recruitment and sampling visits in Scotland. Methods are detailed in reference OP1[57] attached as APPENDIX C.

England & Wales survey: There had been no previous comparable survey in England & Wales. To achieve a prevalence estimate that could be compared with the one for Scotland, it was necessary to visit a minimum of 160 farms; this was done between September 2014 and November 2015.

To recruit farms, a randomly selected subset of suitable farms in England & Wales was acquired from the Rural Payments Agency. These farms were recorded as having a County/Parish/Holding (CPH) number (official designation for all registered cattle keepers) with a premises type of Agricultural Holding, Landless Holding or Temporary Keeper, which contained either a) at least one (non-dairy breed) female aged one year or over, or b) at least one male (any breed) aged one year or over on July 14th 2014.

As in Scotland, all the farmers on this subset list received an initial notification letter, informing them of the study aims and that the survey would start shortly. This provided them with an opportunity to opt out of participation. Those farmers who did not opt out were then phoned, in a randomised order, to arrange a visit to their farm at a convenient time.

The visit procedure was the same as in Scotland. Samples were collected from freshly passed faecal pats, from the group (or groups) of cattle that were closest to being sent to slaughter, with the numbers of samples determined by the same protocol. The samples were packaged and sent by courier to the microbiology laboratory in Inverness.

The recruitment and farm visits in England & Wales were done by members of the ADAS UK Ltd. (now RSK ADAS Ltd.) project team.

Farm management questionnaire: A farm management questionnaire was completed in an electronic format at a face-to-face interview during each sampling visit in both Scotland and in England & Wales. The questionnaire contained information on how each farm was managed, including numbers of different species of animals, approach to buying animals, management of grazing ground and people working on the farm. There were also sections specifically asking about the group sampled for these surveys. These sections included information on whether or not the animals were being kept indoors at the time of sampling, whether their feed or location had changed recently, and whether any of them had been unwell in the weeks before sampling.

Handling data and samples: All the questionnaire information from the Scottish and the English & Welsh farm visits was sent to the project team at the Epidemiology Research Unit of SRUC in Inverness. It was checked for completeness and stored in a secure database. At the microbiology laboratory in Inverness, the same techniques that were used in the previous two Scottish surveys were applied to detect and enumerate *E. coli* O157.

Isolation and identification of *E. coli* O157 was done using immunomagnetic separation (IMS)[59] and enumeration by limiting dilution, performed in duplicate for each sample[60]. Real-Time Polymerase Chain Reaction (PCR) was used to confirm the isolates as *E. coli* O157 and to look for genes encoding Shiga toxin 1 and 2 (ISO/TS, 13136)[57,59,60].

Briefly, within 48 hours of collection, 1g of faeces was suspended in 20ml of buffered peptone water (BPW, ThermoFisher) and incubated at 37°C for six hours. After incubation, 20µl of serogroup O157 IMS beads (ThermoFisher) and 1ml of BPW were concentrated and washed three times. The washed beads were cultured on Sorbitol MacConkey agar supplemented with cefixime and tellurite (CT-SMac, ThermoFisher) and incubated at 37°C overnight. Non sorbitol fermenting colonies were subcultured to Chromocult (Merck) agar and incubated at 37°C overnight. Typical (red) colonies were confirmed as *E. coli* O157 by latex agglutination (ThermoFisher) and PCR.

Methods for prevalence estimation at herd level and pat level: By using statistical modelling methods, the results for the number of herds that were positive for *E. coli* O157 in the current surveys can be used to estimate the number of all similar herds across Scotland and across England & Wales that were likely to be positive for *E. coli* O157 in this time period. This is known as a herd-level prevalence estimate. Using a similar approach, the number of pat samples likely to test positive across all similar herds in Scotland and England & Wales within this time period can be estimated. This is known as a pat-level prevalence estimate.

The surveys described here are samples from a wider population and not a complete census, so a 95% confidence interval (C.I.) must be calculated for these prevalence estimates.

The statistical methods used in this work for both herd-level and pat-level prevalence estimation are described in more detail in Henry et al 2017[57]. The methods used to estimate herd-level prevalence were chosen so that the result for Scotland from this survey could be compared to the results from the two previous surveys in Scotland, and also compared to the current result from England & Wales. In the case of pat-level prevalence, more complex statistical methods were required to enable this comparison between time points within Scotland and between the current estimates for Scotland and for England & Wales. These methods were necessary to account for the particular sampling approach adopted in the second Scottish survey, between 2002 and 2004. These results are reported in the comparative analysis section (3.3).

For each survey, the number of farms on which at least one positive pat in the super-shedder category (here defined as both $>10^3$ and $>10^4$ colony-forming units of *E. coli* O157 per gram of faeces) could be found was also described. When examined using graphical methods, the proportion of pats positive within each sampled group was over-dispersed. This means that either there was greater variability between farms than expected, or the data could be divided into several components.

2.2.3 Results for the two new structured surveys

Herd-level prevalence: *Escherichia coli* O157 could be detected in at least one sample from 26 of the 110 Scottish farms that were visited and 34 of the 160 farms visited in England & Wales. *E. coli* O157 with *stx* genes were found on 25 of these 26 *E. coli* O157 positive Scottish farms and 29 of the 34 *E. coli* O157 positive farms from England & Wales. The outcomes of these prevalence analyses have been published (OP1, Henry *et al* 2017)[57].

The herd-level prevalence estimate for the survey in Scotland was therefore 23.6%, with a 95% confidence interval for this estimate of 16.6-32.5% (Table 2.1). In England & Wales the estimate of the number of similar herds that were likely to have cattle shedding *E. coli* O157 in their faeces during the time period of the survey was 21.3%, with a confidence interval for this estimate of 15.6-28.3% (Table 2.1) There is no statistical difference in this herd-level value between the two current surveys (i.e. between Scotland and England & Wales).

Table 2.1. Herd- and pat-level prevalence estimates for the new British *E. coli* O157 in Cattle Study (BECS) for 2014-2015

Survey	Herd-level prevalence estimate [95% confidence interval]	Pat-level prevalence estimate [95% confidence interval]
BECS (2014-15) Scotland	23.6% [16.6 – 32.5%]	10.6% [6.7 - 16.3%]
BECS (2014-15) England & Wales	21.3% [15.6 – 28.3]	6.9% [4.4 – 10.7%]

Pat-level prevalence estimates: The outcomes of these prevalence analyses have been published[57]. A herd can be positive without all of the samples collected from that herd being positive, so the total number of samples (faecal pats) that were positive in each survey overall was investigated.

A total of 2,763 faecal pats were collected from all the farms in the Scotland survey, of which 647 came from the positive farms. Of these samples, 287 were positive for *E. coli* O157. In Scotland, 10.6% of the pat samples collected across all farms were estimated to contain *E. coli* O157, with a 95% confidence interval for this estimate of 6.7-16.3% (Table 2.1).

In the England & Wales survey, 2,913 faecal pats were collected from all the farms visited, of which 778 were collected from the positive farms. Of these pat samples, 234 were positive for *E. coli* O157. In England & Wales, 6.9% of the samples collected across all farms were estimated to contain *E. coli* O157, with a confidence interval for this estimate of 4.4-10.7%. (Table 2.1).

There is no statistical difference in this pat-level prevalence estimate between the two current surveys.

Cut-offs for classifying super-shedder samples: There has been some debate about the quantity of *E. coli* O157 bacteria a sample should contain for it to be classified as a super-shedder sample[6,30,39]. For the prevalence study, two different cut-off values were used: 10^3 (i.e. 1000) and 10^4 (i.e. 10000) colony-forming units per gram (cfu/g) of sample. Samples with a bacterial count of over 10^3 cfu/g were found on 10 of the 26 positive farms in Scotland and on 13 of the 34 positive farms in England & Wales. Samples with a bacterial count of over 10^4 cfu/g were found on nine of the 26 positive farms in Scotland and on seven of the 34 positive farms in England & Wales.

2.3 Comparative analyses of survey data (DO 2.1.1 cont.)

This section of the report focuses on the comparison of prevalences and strain composition within the current survey and across the historical surveys conducted in

Scotland. The comparative analyses were performed at three different levels (Table 2.2) examining the prevalence (in cattle) and incidence (in human) as well as differences in strain composition using Phage Type and *stx* subtyping. Overall differences were examined in addition to differences associated with season, Animal Health District (AHD, Fig. 2.1) and shedding level. Seasons were defined as winter, comprising December, January, and February; spring, comprising March, April and May; summer, comprising June, July and August; and autumn, comprising September, October and November. Six regions, based on Veterinary Animal Health Districts (AHDs) were defined: 1 = Islands; 2 = Highland; 3 = North East; 4 = Central; 5 = South East; 6 = South West. AHDs are no longer used within Scotland, however, they were retained within this study to maintain continuity. Shedding was classified as low-level (low, $<10^3$), moderate-level (moderate, 10^3-10^4) and high-level (high, super-shedding, $>10^4$) based on the number of cfu/g of faeces. This was done to capture the historical and more modern definitions of super-shedding. Whole genome sequencing results are presented in a later section (2.5) led by Public Health England (PHE).

Figure 2.1. Animal Health Divisions (AHDs) based on Veterinary AHDs: 1 = Highland; 2 = North East; 3 = Central; 4 = South West; 5 = South East; 6 = Islands.

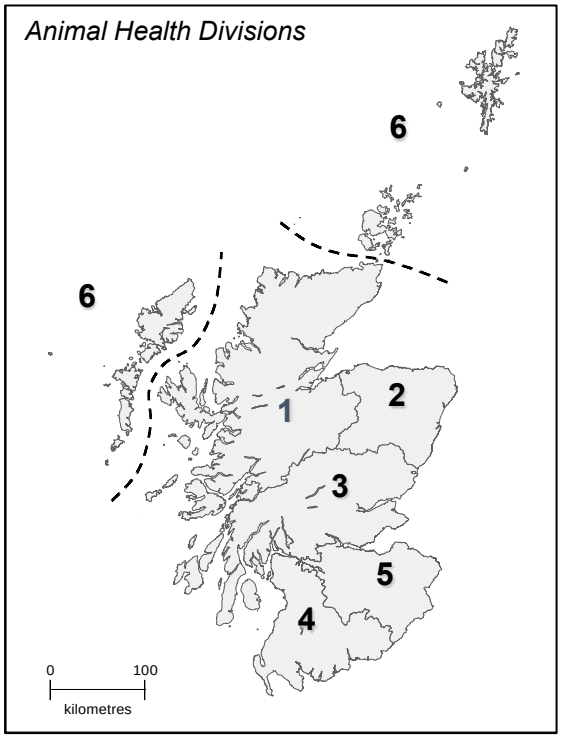


Table 2.2. Level of comparison including a description of the comparison that will be performed

Level	Description
1	Cattle only, within Scotland. A. prevalence*; B. strain composition
2	Cattle only, Scotland versus England & Wales; A. prevalence*; B. strain composition
3	Cattle versus human, Scotland versus England & Wales. A. prevalence; B. strain composition

*Prevalence calculations for Survey 3 (BECS: 2014-2015) are provided above and in Henry *et al* 2017[57].

2.3.1 Level 1A: Comparison of current Scottish cattle E. coli O157 prevalence estimates to historical prevalence values

Between March 1998 and September 2015, three cross sectional surveys were conducted in Scotland (Fig. 2.2). The field sampling methodologies for all surveys have been published[40,57,58], however, a brief outline is given below and in Fig. 2.2 highlighting how the survey methodology differed between the three surveys.

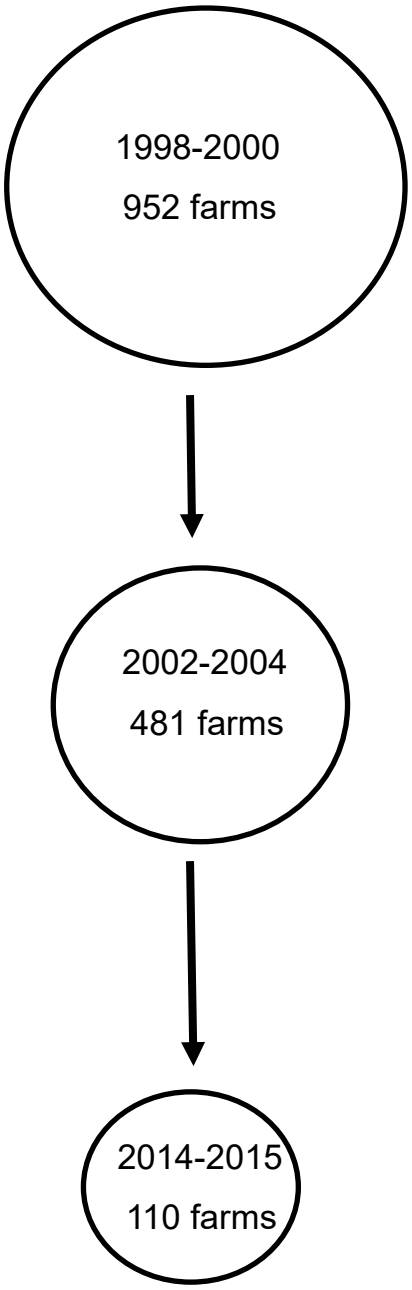
All surveys preferentially sampled cattle groups composed only of store (i.e. weaned cattle before finishing for slaughter) or finishing cattle closest to sale or slaughter. If such groups did not exist, one or more mixed groups with store or finishing cattle closest to sale or slaughter were sampled. From each group, fresh faecal pats were sampled. The number of pats tested in each group was determined from the number of cattle in the group using a prescribed sampling schedule.

For Survey 1, sufficient numbers of faecal pats were tested to ensure prospectively an 80% chance of sampling at least one positive pat if there was a shedding prevalence of at least 2% within the group[58]. Farms were randomly chosen (n=952) from a selected population of over 3,111 farms, which represent about 20% of the 12,000 agricultural holdings in Scotland that have cattle, including some dairy farms if they also keep animals destined for the food chain.

Based on results from Survey 1, in Survey 2, it was assumed that, on average, 8% of the animals in positive groups would be shedding *E. coli* O157, with shedding distributed as seen in Survey 1[40]. For each group in Survey 2, sufficient fresh pat samples were taken to ensure prospectively a mean 90% probability of detecting shedding of *E. coli* O157 if at least one shedding animal was indeed present. Changes in sampling strategy between the two surveys had a negligible effect on the power to identify positive farms[40]. Instead of randomly sampling farms, Survey 2 used a stratified sampling plan derived from the Survey 1 cohort to select farms to

sample[40]. Farms were selected randomly from this set and then farms that were close in distance were sampled on the same or concurrent days. Thirty-four farms were added in the Highlands and Islands to ensure representation in those districts. In survey 3, 110 farms were sampled, drawn primarily (102/110) from the farms that were sampled in both Survey 1 and Survey 2[57].

The sampling of the same farms across the surveys and the comparable methodology allow the use of these separate surveys as three cross-sectional time points for our analysis. The aim was to compare the overall herd-level and pat-level prevalences across all 3 surveys as well as examine seasonal, spatial and phage type differences. In addition, we tested an additional hypothesis using the same statistical models to determine if there was an association between a farm's current status and its status in a previous survey.



Survey 1 (SEERAD): 1998-2000.
Sample design: Stratified random sampling of 952 Scottish farms with beef finishing and store cattle from a sampling frame of 3,111 farms with cattle, randomly selected from 1997 Scottish Agricultural and Horticultural Census data.
Power: 80% probability of detecting at least 1 positive assuming 2% shedding within group.
Spatial distribution: National; six regions: five Animal Health Divisions plus the Northern and Western Isles forming a separate region
Typing data: Phage Typing

Survey 2 (IPRAVE): 2002-2004.
Sample design: Cluster sampling of 481 Scottish farms with beef finishing and store cattle from 925* farms sampled in Survey 1 plus 34** new farms in the Highlands and Islands.
Power: 90% probability of detecting at least 1 positive assuming 8% shedding within group.
Spatial distribution: National; six regions: five Animal Health Divisions plus the Northern and Western Isles forming a separate region
Typing data: Phage Typing, Pulse-Field Gel Electrophoresis, *stx* subtyping[¥], Whole Genome Sequencing

Survey 3 (BECS): 2014-2015.
Sample design: Random sampling of 110 Scottish farms with beef finishing and store cattle sampled randomly from the farms sampled in Survey 1 and Survey 2.
Power: 96% confidence; tolerance ± 0.169[Ψ].
Spatial distribution: National; six regions: five Animal Health Divisions plus the Northern and Western Isles forming a separate region
Typing data: Phage Typing, *stx* subtyping, Whole Genome Sequencing

Figure 2.2. Historical sampling of *E. coli* O157 in Scotland. Schematic diagram describing the sampling design and the typing information available for each of the three cross-sectional surveys conducted in Scotland (1998-2015). * 925 of the 952 (97%) farms agreed to be involved in Survey 2. ** Farms were added to ensure good representation for Highlands and Islands districts. ¥, Only 29% of the *E. coli* O157 isolated were *stx* subtyped. Ψ, the BECS survey was powered so as to achieve the same confidence interval (tolerance) as achieved in the IPRAVE survey.

Herd-level Prevalence: There was no significant difference in the herd-level prevalence across surveys (Table 2.3; Fig. 2.3A). In Scotland, roughly 20% of the farms are likely to have cattle shedding *E. coli* O157. This figure is consistent across Scotland as there were no differences within geographical regions. For this study we have continued to use Animal Health Districts (AHD, see Fig. 2.1) (p=0.181) (Table 2.3 and Figure 2.5), which were the regional areas used for analyses in Surveys 1 & 2. We have retained these regions for consistency and to allow spatial comparisons within and across surveys to be performed.

There were significant seasonal differences identified (p=0.001) with summer and autumn prevalences significantly higher than winter (Fig. 2.3B). This pattern was the same across the 3 surveys.

Pat-level Prevalence: There were significant differences in the pat-level prevalence across surveys (p<0.001) with both Survey 1 and Survey 3 being significantly higher than Survey 2 (Table 2.4; Fig. 2.2C). Although there were no spatial differences (AHD: p=0.234) (Table 2.4) there were seasonal differences (Figure 2.3D). There appeared to be a different seasonal pattern across surveys, with the pat-level prevalences being the highest in the winter during Survey 3 yet the lowest in Survey 1 and Survey 2 although the power was not sufficient to tease out this relationship. This will be investigated in future work using a spatial-temporal risk factor model.

Is a farm's past status a significant predictor of current status? Of the 102 farms that were sampled across all 3 cattle surveys that are included in this analysis no farms were positive across all 3 surveys; 48 were not positive for any survey; and 54 were positive in at least 1 survey. The latter result was not associated with being PT21/28 (exact χ^2: 1.57, p=0.298). Being positive in the current survey is not associated with being positive for any survey in the past (exact χ^2 test: 0.220, p=0.807).

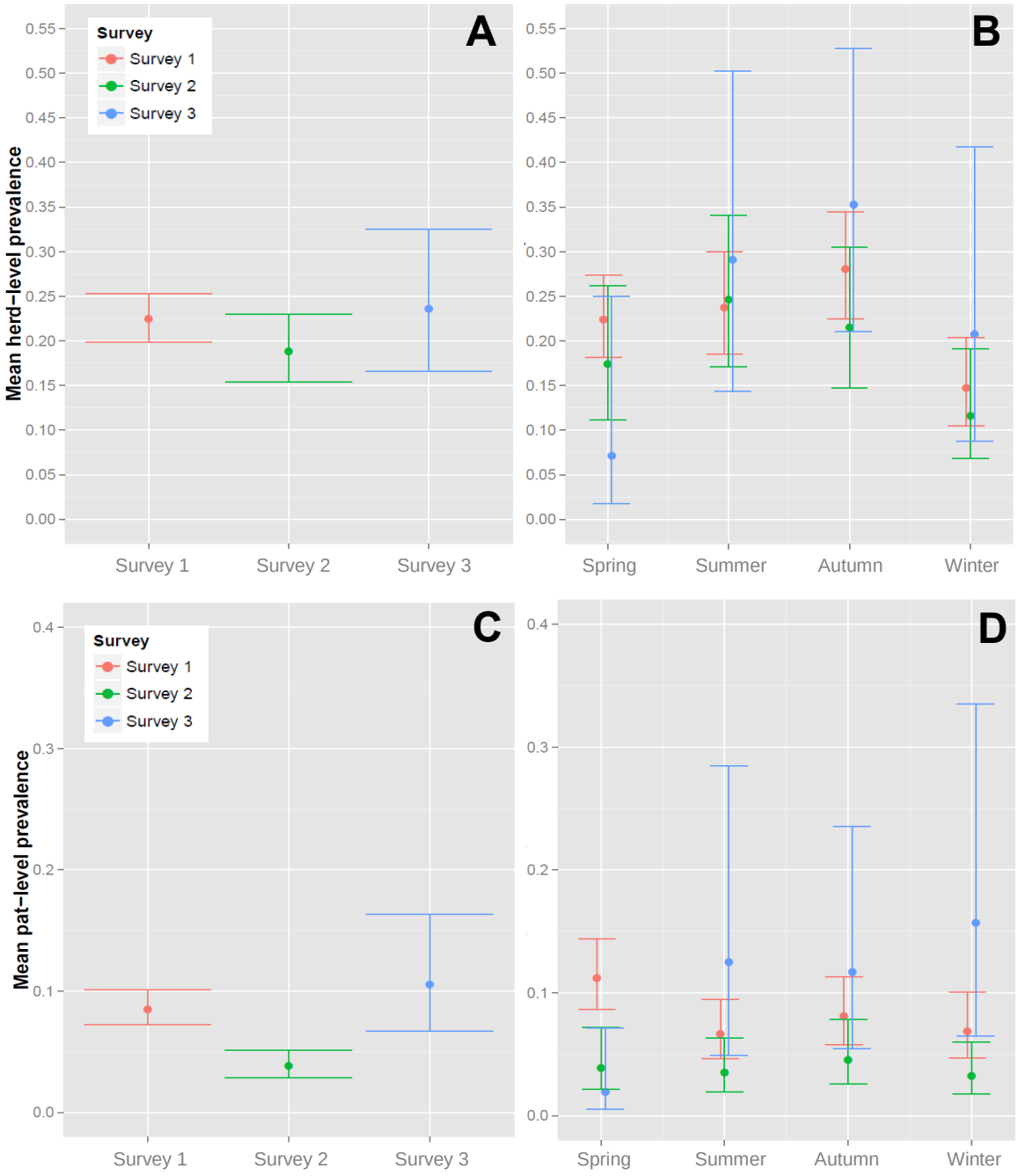


Figure 2.3. Cattle Prevalence. Mean prevalence for Scottish cattle from Survey 1 (SEERAD: 1998-2000), Survey 2 (IPRAVE: 2002-2004) and Survey 3 (BECS: 2014-2015) overall herd-level (A) and pat-level (C) prevalence as well as prevalence by season for herd-level (B) and pat-level (D).

Table 2.3: Mean farm-level prevalence of bovine *E. coli* O157 shedding for: Survey 1 (SEERAD: March 1998 – May 2000, n=952 farms), Survey 2 (IPRAVE: February 2002 – February 2004, n=481 farms) and Survey 3 (BECS: September 2014 – September 2015, n=110 farms) surveys. Season: spring (March-May), summer (June-August), autumn (September-November), winter (December-February); AHD, Animal Health District.

Category	Mean Prevalence (lower, upper 95% confidence limits)		
	Survey 1 (SEERAD)	**Survey 2 (IPRAVE)**	**Survey 3 (BECS)**
All categories	22.5 (19.9, 25.3)	18.9 (15.4, 23.0)	23.6 (16.6, 32.5)
By season			
Spring	22.4 (18.2, 27.4)	17.4 (11.1, 26.2)	7.1 (1.7, 25.0)
Summer	23.8 (18.5, 30.0)	24.6 (17.1, 34.1)	29.2 (14.4, 50.2)
Autumn	28.1 (22.5, 34.5)	21.6 (14.7, 30.5)	35.3 (21.0, 52.8)
Winter	14.7 (10.4, 20.4)	11.6 (6.8, 19.1)	20.8 (8.8, 41.8)
By AHD			
Highland	16.8 (10.5, 25.9)	17.9 (10.6, 0.28.8)	24.9 (10.4, 48.5)
North East	25.7 (20.2, 32.1)	18.3 (10.8, 29.3)	26.3 (11.1, 50.6)
Central	24.8 (19.0, 31.6)	18.7 (10.9, 30.2)	20.0 (6.4, 47.8)
South West	20.0 (15.4, 25.5)	24.5 (15.6, 36.3)	17.5 (6.6, 38.9)
South East	25.2 (18.5, 33.4)	23.7 (15.0, 35.4)	29.2 (12.3, 54.9)
Islands	18.0 (11.0, 27.9)	9.9 (4.9, 19.0)	24.9 (9.3, 51.6)

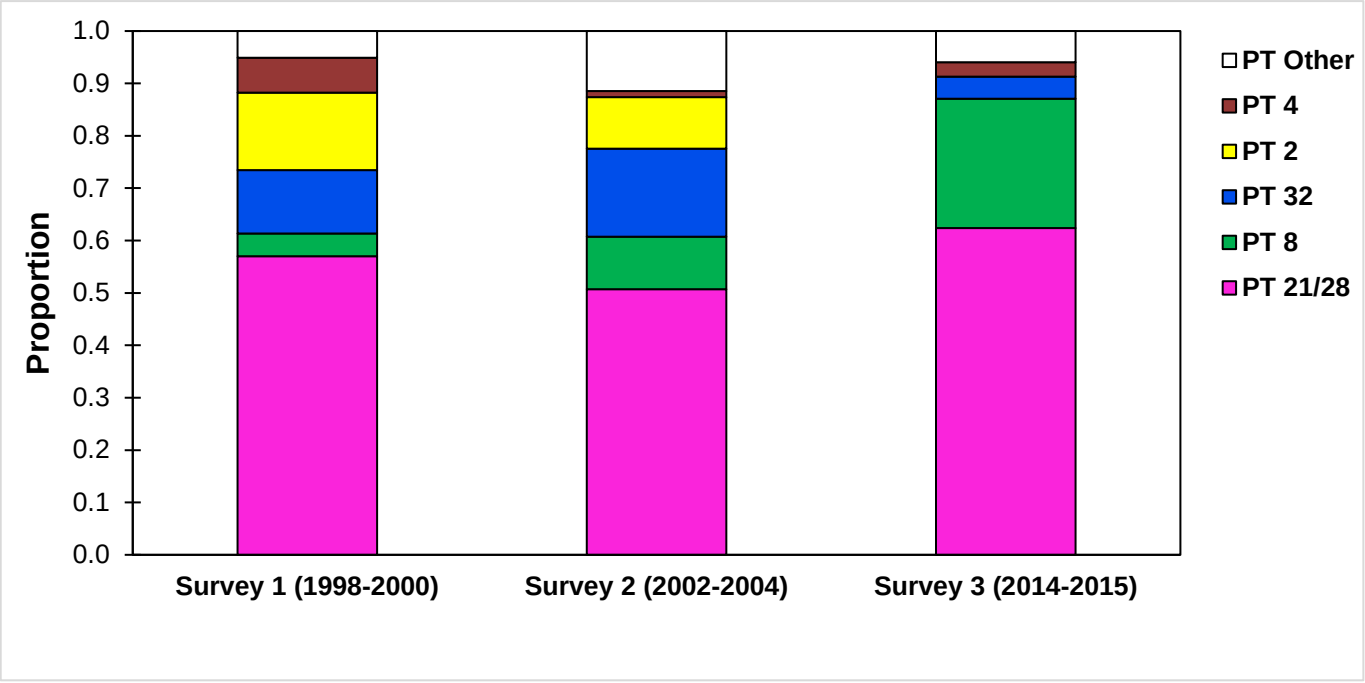
Table 2.4: Mean pat-level prevalence of bovine *E. coli* O157 shedding for: Survey 1 (SEERAD: March 1998 – May 2000, n=952 farms), Survey 2 (IPRAVE: February 2002 – February 2004, n=481 farms) and Survey 3 (BECS: September 2014 – September 2015, n=110 farms) surveys. Season: spring (March-May), summer (June-August), autumn (September-November), winter (December-February); AHD, Animal Health District.

Category	**Mean Prevalence (lower, upper 95% confidence limits)**		
	Survey 1 (SEERAD)	**Survey 2 (IPRAVE)**	**Survey 3 (BECS)**
All categories	8.5 (7.2, 10.1)	3.8 (2.9, 5.1)	10.6 (6.7, 16.3)
By season			
Spring	11.2 (8.7,14.4)	3.9 (2.1,7.2)	2.0 (0.5, 7.2)
Summer	6.7 (4.7, 9.5)	3.5 (2.0, 6.3)	12.5 (4.9, 28.4)
Autumn	8.1 (5.8, 11.3)	4.6 (2.6, 7.9)	11.7 (5.4, 23.5)
Winter	6.9 (4.7, 10.1)	3.3 (1.8, 6.0)	15.7 (6.5, 33.5)
By AHD			
Highland	9.8 (5.9, 15.8)	2.0 (0.90, 4.5)	9.1 (3.0, 24.5)
North East	9.7 (7.0, 13.2)	2.7 (1.3, 5.7)	15.5 (5.7, 35.8)
Central	9.6 (6.8, 13.6)	3.9 (1.9, 7.7)	15.6 (5.0, 39.4)
South West	5.8 (4.0, 8.4)	7.4 (4.1, 13.0)	10.0 (3.8, 23.8)
South East	10.6 (7.1, 15.5)	4.5 (2.3, 9.1)	5.8 (1.5, 20.3)
Islands	5.7 (2.9, 11.1)	3.4 (1.6, 7.0)	7.9 (2.0, 26.5)

2.3.2. Level 1B: Comparison of current strain composition of Scottish cattle E. coli O157 with historical values (DO 2.1.4)

What follows is a comparison of the cattle *E. coli* O157 strain composition within Scotland (Level 1B, Table 2.2) and between Scotland and England & Wales (Level 2B, Table 2.2). For the purpose of this analysis, shedding was categorised as low ($<10^3$ cfu/g of faeces); moderate ($10^3 – 10^4$ cfu/g of faeces) and high ($>10^4$ cfu/g of faeces). Strain composition was examined using PT and *stx* subtype. PT was the only typing method that was performed across all 3 surveys (Fig 2.2). PT maps well onto the core genome and has been used classically in identifying related strains, for example in outbreak investigations. *Stx* subtyping is only available for survey 3 and a subset of strains from Survey 2.

The *E. coli* O157 shed from Scottish cattle across the three surveys is dominated by 5 main PTs that comprised >90% of all PTs found in Scotland: PT2, P4, PT8, PT32 and PT21/28. Other PTs include PT14, PT24, PT31, PT33, PT34, PT49 and PT54 and represent <10% of all the typeable strains. Since the first survey there has been a loss of PT2, a decrease of PT32 (through the loss of PT32 *stx*2(a+c)) and an increase of PT8 (Fig. 2.4). The proportion of isolates that were PT21/28 in Survey 3 was higher but not significantly different from Survey 1 (p=0.098), although was significantly higher than Survey 2 (p=0.002) (Fig. 2.4, upper figure). PT21/28 is primarily *stx*2(a+c) hence the high percentages of *stx*2(a+c) in Survey 2 and Survey 3 (Fig. 2.4 lower figure), although the PT21/28 recovered from one farm in the Highland AHD was *stx*1+*stx*2(a+c). There has been an increase in *stx*2a alone between Survey 2 and Survey 3 (p=0.0408), however, there has been no increase in the overall amount of *stx*2a (i.e. *stx*2(a+c)) (p=0.2488). In Survey 2 only 2/151 (1.3%) isolates that were *stx*2 subtyped were *stx*2a alone as opposed to 15/279 (5%) in Survey 3 (Fig. 2.4 lower figure).



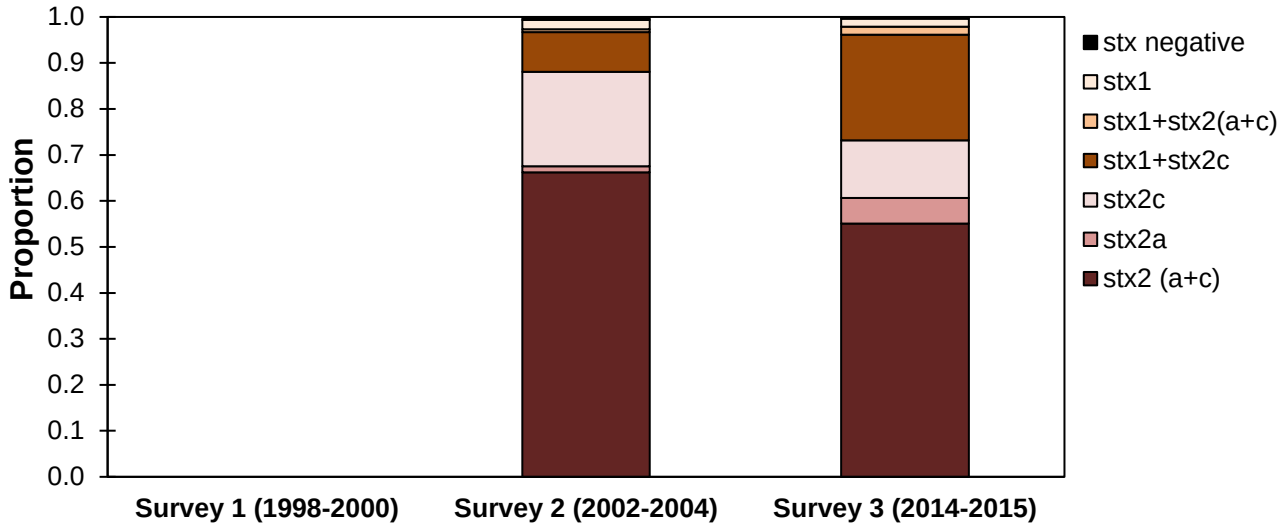


Figure 2.4. Stacked bar graphs representing the strain composition of cattle *E. coli* O157 collected in Scotland: Survey 1 (SEERAD: 1998-2000), Survey 2 (IPRAVE: 2002-2004) and Survey 3 (BECS: 2014-2015). Strain composition is characterised by Phage Type (PT) (upper figure) and *stx* subtype (lower figure).

Fig. 2.5 shows the proportion of farms that were positive in each of the 6 AHDs in Scotland across the 3 surveys. Positive farms were divided into the proportion of farms that were PT21/28 (pink) and the proportion that were not PT21/28 (white). Analysis of the proportion of PT21/28 farms across AHD stratified by survey was significant (Cochran-Mantel-Haenszel Test, p=0.042) showing that there were differences across the surveys with respect to the proportion of farms in AHDs that were PT21/28 (Survey 1 v Survey 2: p=0.0419; Survey 2 v Survey 3: p=0.0042). There were no differences across AHDs in terms of the proportion of farms that were PT21/28 positive in Survey 1 (Fisher-Freeman-Halton test: p=0.1733) but there were

in Survey 2 (p=0.0265) and Survey 3 (p=0.011). The proportion of PT21/28 on positive farms in the first survey ranged from 43%-74%, in Survey 2 from 30% - 88% and in Survey 3 from 0% to 100%. In Survey 3 positive farms in the North East, Highlands and Islands were primarily PT21/28 (>80%). Strains from farms from the Highlands and Islands were entirely PT21/28 (100%) (Fig. 2.5).

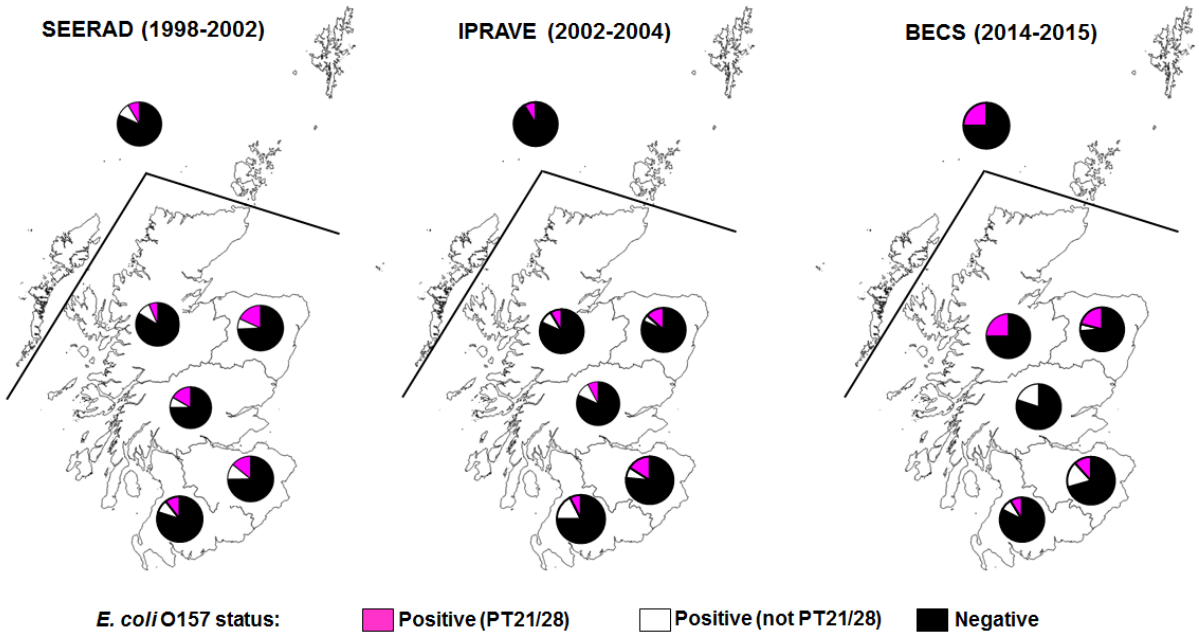


Figure 2.5. The proportion of farms positive in each of the 6 Animal Health Districts (AHDs) in Scotland for the three national cross-sectional surveys. Positive farms were separated into the proportion that were PT21/28 (pink) and the proportion that were not PT21/28 (white). The proportion of farms that were negative is shown in black. AHDs consist of Highland, Islands, North East, Central, South East and South West. APPENDIX D contains the data associated with this figure.

Data on the counts of *E. coli* O157 (in cfu/g of faeces) were only performed for Survey 2 and Survey 3 (Fig. 2.2). There was an increasingly higher proportion of high-level shedders that were PT21/28 for both Survey 2 (p=0.001) and Survey 3 (p=0.013) (Fig. 2.6-top panel). There were, however, differences between the surveys (singly ordered Cochran-Mantel-Haenszel test: p<0.001). In Survey 2, compared to low level shedders ($<10^3$), PT21/28 was significantly associated with both moderate (10^3-10^4) (p=0.047) and high-level, super-shedding ($>10^4$) (p=0.003), however, in Survey 3, compared to low level shedders ($<10^3$), PT21/28 was only associated with moderate shedding (10^3-10^4) (p=0.018) and not high level / super-shedding ($>10^4$) (p=0.252). For Survey 2, there was no difference in the proportion of *stx*2a (alone or with *stx*2c) between low and moderate shedders (p=1.00) but high-level shedders had significantly higher proportions of *stx*2a than low (p=0.002) and moderate shedders (p=0.03) (Fig. 2.6-bottom panel). For Survey 3, there were a

higher proportion of moderate-level shedders than low-level shedders that were *stx*2a (alone or with *stx*2c) (p=0.018) but the proportion of high-level shedders that were *stx*2a was not significantly higher than the moderate level shedders (p=0.7716).

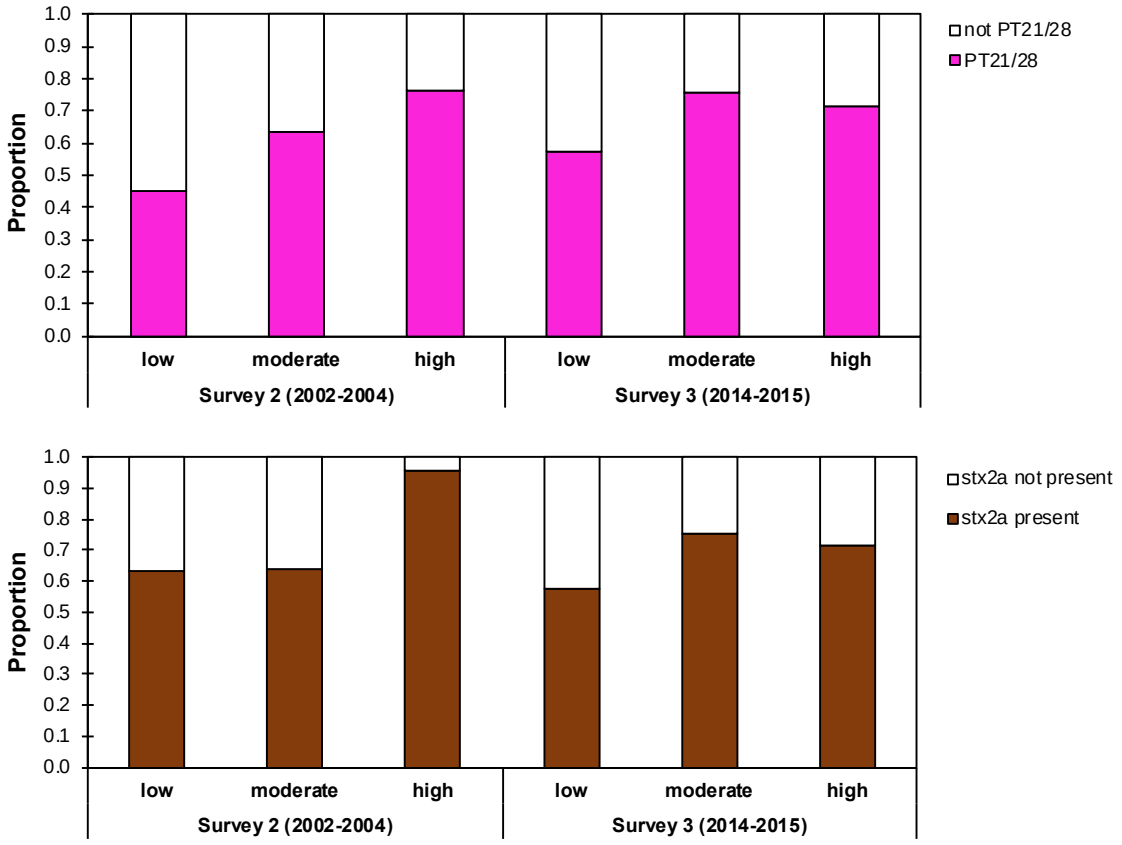


Figure 2.6. Stacked bar graphs representing the strain composition for the different shedding levels of cattle *E. coli* O157 collected in Scotland: Survey 1 (SEERAD: 1998-2000), Survey 2 (IPRAVE: 2002-2004) and Survey 3 (BECS: 2014-2015). (A-top panel) Strain composition is characterised by Phage Type (PT: PT21/28 and not PT21/28) and (B-bottom panel) *stx* subtype (*stx*2a present and *stx*2a not present). Low-level shedding (low, $<10^3$); moderate-level shedding (moderate, 10^3-10^4); and high-level (high, super-shedding $>10^4$).

2.3.3. Level 2: Survey 3, Cattle only, Scotland versus England & Wales; A. prevalence; B. Strain composition (DO 2.1.4)

Comparison of the prevalences of *E. coli* O157 in cattle for Survey 3 are presented above (Table 2.1 and Henry et al. 2017[57]). Results showed that there were no statistically significant differences in either the herd-level or pat-level prevalences between England & Wales and Scotland. Although there were no differences in the prevalence of cattle shedding, the strain composition (defined by PT and *stx* subtype) were significantly different (Fig 2.7). On farms in Scotland the dominant PT

was PT21/28, found on 17/26 (65%) positive farms. In England & Wales, however, PT 21/28 was observed on only 3/34 (9%) positive farms. These farms were located in the North East, West Midlands and Wales (Fig. 2.7).

Farms in Scotland and England & Wales were equally likely to have cattle shedding at moderate ($>10^3$) and high ($>10^4$) levels (Henry et al., 2017). However, the distribution of shedding levels within pats is different between Scotland and England & Wales (E&W) (chi-square 13.78; p=0.0013) (Fig 2.8 upper figure). Both countries had cattle that shed predominately (>70%) at low levels, however, the proportion of low level shedding cattle was higher in E&W than Scotland. Scottish cattle had higher levels of moderate and high level shedding. In Scotland, the majority of moderate and high level shedding was associated with PT21/28. No PT21/28 were associated with moderate (10^3-10^4) or high ($>10^4$) shedding in E&W. Only 9 isolates of PT21/28 on 3 farms were found in E&W (two in England and one in Wales) none of which had *E. coli* counts $>10^3$. To ascertain if this lack of association was related to the small sample size, bootstrapping was performed where a sample of size 9 was repeatedly (n=10,000) taken from the Scottish sample of PT21/28s. The results show that an observation of no animals shedding $>10^3$ is possible in 3.5% of samples with a size of 9. There is no dominant PT in cattle in E&W and there is no one strain that is associated with high shedding (Fig 2.8 lower figure).

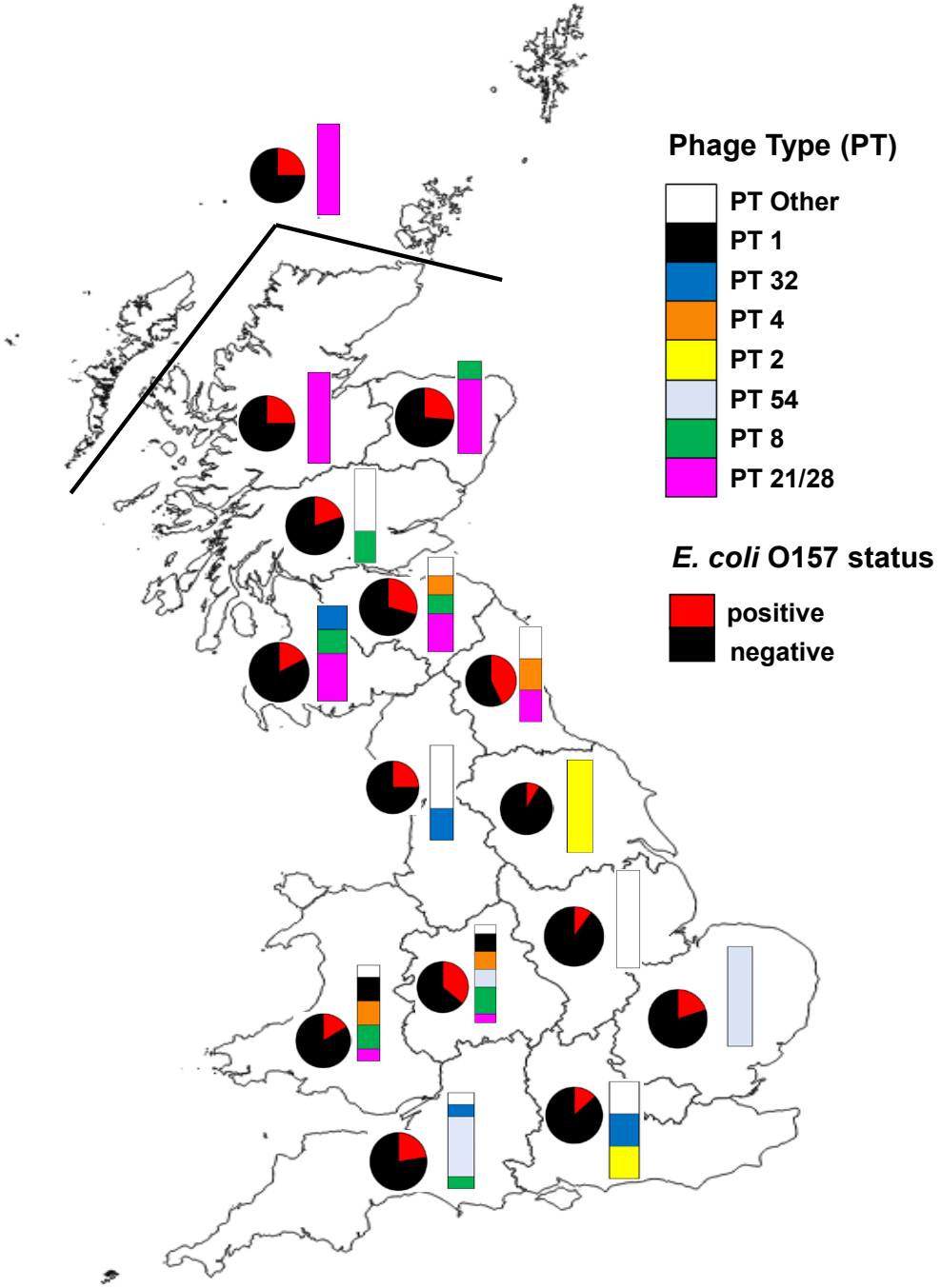


Figure 2.7 Map of the UK divided into regions. In Scotland, 6 Animal Health Districts (AHDs) include: Highland, Islands, North East, Central, South East and South West. In England & Wales spatial regions were defined using the Nomenclature of Units for Territorial Statistics (NUTS). NUTS 1 regions for England & Wales include: North East, North West, Yorkshire, East Midlands, West Midlands, East of England, London (no data), South East, South West, Wales pie charts show the proportion of farms in the areas that were positive (red) and negative (black). Stacked bars show the proportion of positive samples that were different PTs: PT1, PT32, PT4, PT2, PT54, PT8, PT21/28 and PT Other (other includes the following PTs: 14, 31, 34, RDNC). APPENDIX D contains the data associated with this figure.

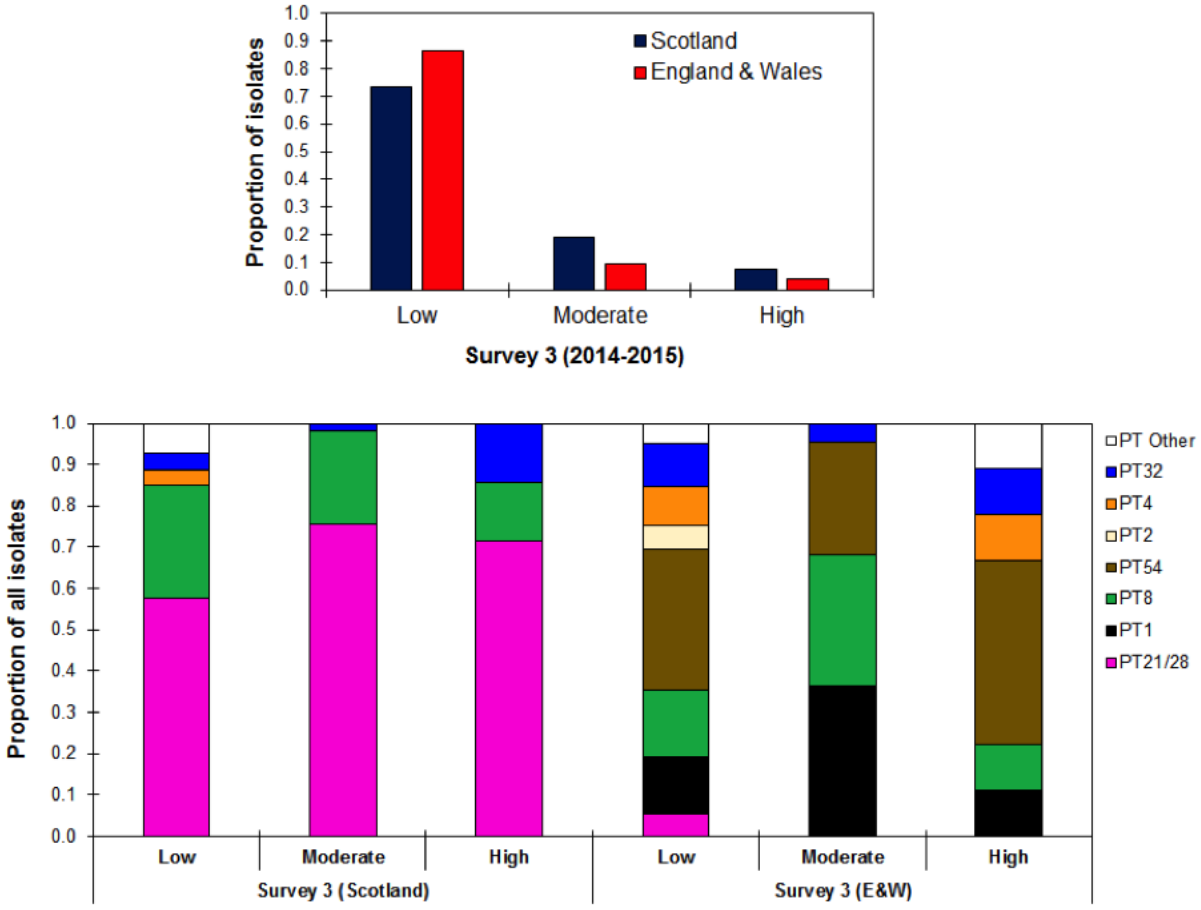


Figure 2.8. Proportion of all isolates (top panel) and the PT strain composition (bottom panel) for the different shedding levels of cattle *E. coli* O157 collected in Scotland and England & Wales (E&W) during Survey 3 (BECS: 2014-2015). Low shedding ($<10^3$); moderate shedding (10^3-10^4); and high-level or super-shedding ($>10^4$). Scotland is twice as likely to have a sample with counts $>10^3$ (chi-square: 13.78; p=0.0003; odds (95% CI): 2.36 (1.46-3.89)) and $>10^4$ (chi-square: 3.13; p=0.090; odds (95% CI): 2.04 (0.871-5.15)) although only shedding $>10^3$ was statistically significant.

2.3.4. Level 3: Cattle versus human, Scotland versus England & Wales. A. prevalence; B. strain composition (DO 2.1.4)

For cattle, there was no significant difference between the herd-level and pat-level prevalences between Scotland and England & Wales (Table 2.1). For human clinical cases, there is an overall decreasing trend in the rate of human clinical cases of 0.58% per year in Great Britain between 1998 and 2017 but this is not statistically significant ($p=0.294$) as shown in Fig. 2.9. Although the pattern of change with time is similar in Scotland and England & Wales, the rate of human clinical cases is significantly lower in England & Wales when compared to Scotland ($p<0.001$) (Table 2.5, Fig. 2.9). Within the human cases, PT21/28 and PT8 are now the predominant PTs observed (Fig. 2.10). For this comparison, cattle data from Survey 3 were used along with human reported clinical cases from the same time frame. In England & Wales all travel-related illness and outbreak cases were removed. For Scotland, known travel-related cases have been removed but these results should be deemed preliminary until the data is finalised (Fig. 2.11).

In Scotland the proportion of isolates sampled from cattle that have a *stx*2a gene (either alone or with *stx*2c) is very similar to the proportion with *stx*2a from human isolates (both ~60% of the isolates). However, for England & Wales the proportion of samples from cattle with *stx*2a is substantially lower (~15%) compared to ~50% in human cases. In Scotland the high proportion of *stx2a* observed in both cattle and human isolates is associated with the presence of PT21/28 (which are all *stx*2a positive). In England & Wales, although PT21/28 comprises a large proportion of human clinical cases it was not observed in a high proportion in the cattle samples. The disparity in the England & Wales data between cattle samples and strains associated with human infection is not currently understood. It is possible that the discrepancy is a result of the sampling effort. The survey was powered to determine the overall prevalence of *E. coli* O157 and not that of each strain. However, it could reflect that more *stx*2a+ and PT21/28 *E. coli* O157 infections in England & Wales are coming from environmental, animal or food sources distinct from the local cattle reservoirs.

Table 2.5. Values for cattle prevalence (herd-level and pat-level) from Survey 3 for Scotland and England & Wales along with the rate per 100,000 population of reported human clinical cases for Scotland and England & Wales for the years 2014-2017.

	England & Wales		Scotland	
	Cattle	Human	Cattle	Human
Herd-level prevalence* 2014-15	21.3		23.6	
Pat-level prevalence* 2014-15	6.9		10.6	
Rate per 100,000 population**				
2014		1.5		4.9
2015		1.1		3.2
2016		1.2		3.4
2017		1.0		3.1

*Cattle prevalence estimates from the BECS (Survey 3, 2014-2015)[57]

**Published data on number of reported clinical cases from Health Protection Scotland (Scotland) and Public Health England (England & Wales) for 2014-17

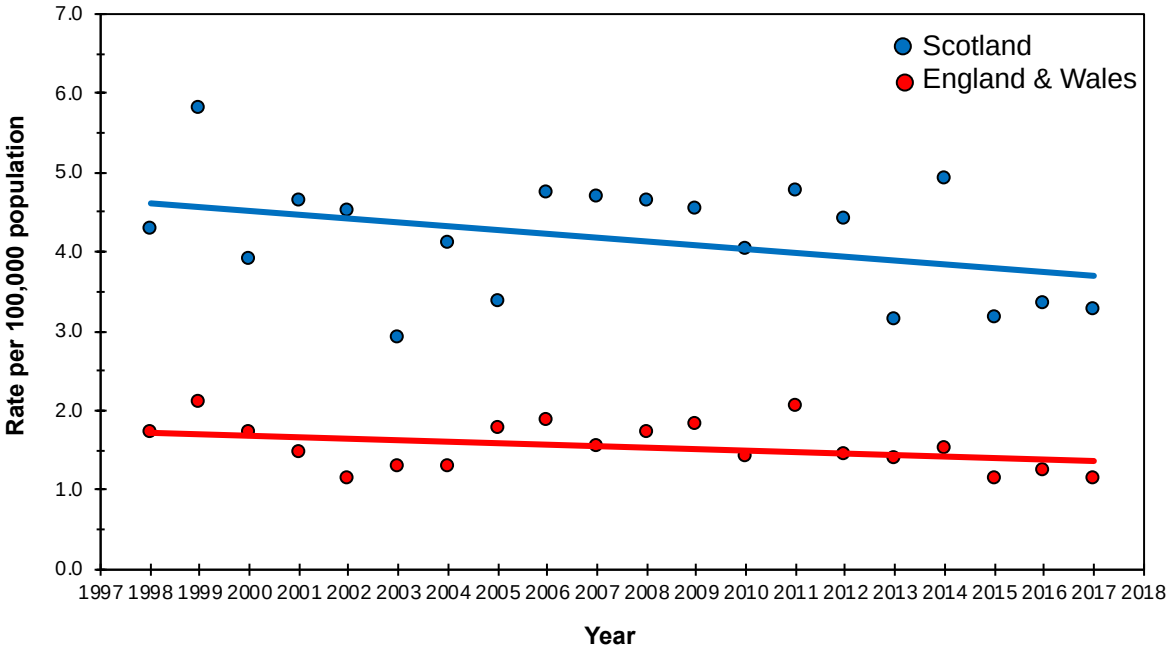


Figure 2.9. **Rate per 100,000 population of reported human clinical cases of *E. coli* O157 for Scotland and England & Wales from 1998-2017**. Lines represent the linear trend. Poisson regression of analysis with offset for population size shows significant differences between Scotland and England & Wales ($p<0.001$), however, no difference across time ($p=0.2306$).

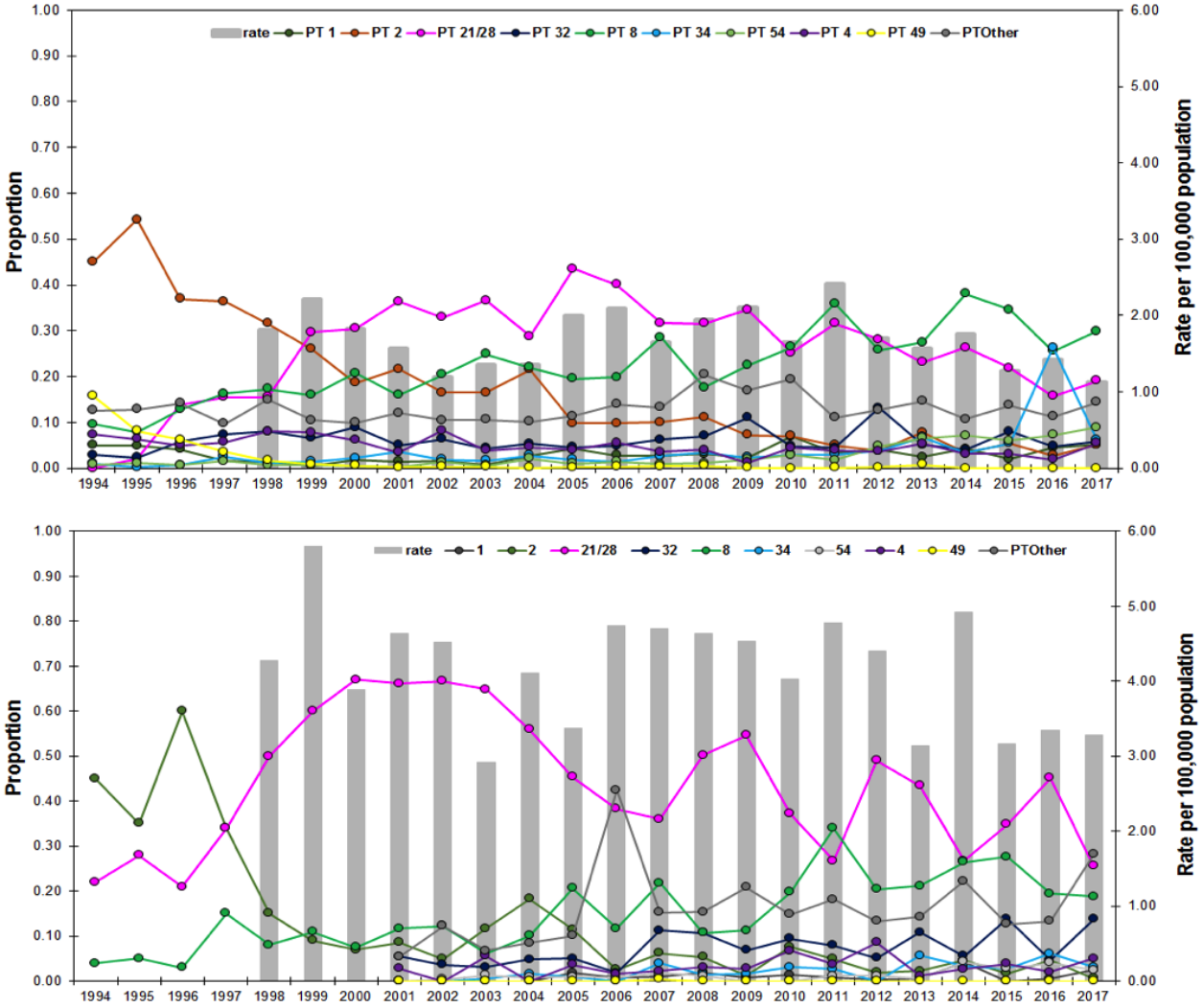


Figure 2.10. Temporal trends in the rate (per 100,000 population) and Phage Type (PT) distribution for England & Wales (upper figure) and Scotland (lower figure). Includes all samples submitted to SERL and PHE over a given calendar year (1994-2017) thus travel-related and outbreak cases are included.

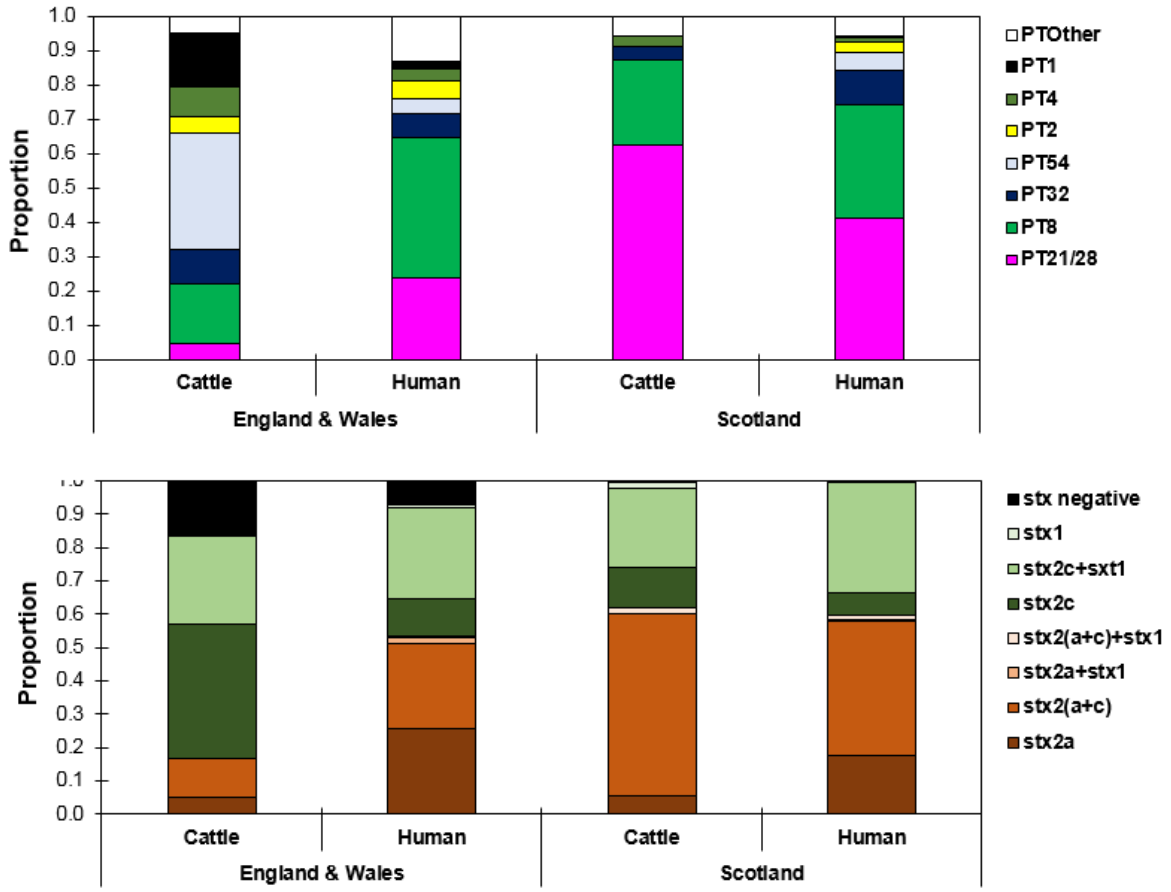


Figure 2.11. Stacked bar graphs representing the strain composition of cattle isolates and human clinical reported cases in Scotland and England & Wales during Survey 3. Strain composition is described by Phage Type (PT) (upper figure) and *stx* subtype (lower figure). All travel-related cases have been removed from data from England & Wales. In Scotland known travel-related cases have been removed but these results should be deemed preliminary until the data is finalised.

2.4. Risk factors at farm level (final part of DO 2.1.4)

Descriptions of the farms: In the England & Wales survey, questionnaire responses were completed for 159 of the 160 farms. In Scotland all 110 were completed.

Farm Type: The main cattle management type on sampled farms across both surveys was suckler beef (73/110 in Scotland and 109/159 in England & Wales) – a system in which calves are raised for beef and kept with their mothers and the main herd until around six to nine months of age, at which point they are sold on or managed separately. The remaining herds were dairy systems (producing milk), specialist finisher systems (where older calves are kept in groups and fattened until

they are at a suitable weight for slaughter) or a category described as “other”, which was most frequently a combined suckler beef and specialist finisher system – again, this fits with the expected herd types for this study.

Herd Size**:** In Scotland, study herds that were dairy systems tended to have more cattle than suckler beef systems and specialist finisher systems. Both dairy and specialist finisher systems tended to have more cattle between 12 and 30 months of age than suckler beef systems. In England & Wales, dairy systems tended to have more cattle than suckler beef systems. Specialist finisher systems tended to have more cattle between 12 and 30 months of age than suckler beef systems. Cattle in this age group which are not intended for breeding are typically being kept for slaughter.

In Scotland, the average total number of cattle on the farm on the day of sampling was 176. The average number of cattle in the sampled group was 17. There was no difference between positive and negative farms in terms of the average number of cattle in the sampled group or the average total number of cattle on the farm.

In England & Wales, the average total number of cattle on the farm on the day of sampling was 85, with an average of 14 cattle in the sampled group. Differences were seen between positive and negative farms in terms of the average total number of cattle on the farm and the average number of cattle in the sampled group. Positive farms tended to have more cattle overall and the sampled groups were larger than on negative farms.

Age of sampled animals: The median ages of the youngest and oldest animals in the sampled groups were 15 and 22 months in Scotland and 14 and 20 months in England & Wales.

Other management factors: The majority of sampled groups were housed at the time of sampling in both Scotland (83/110, 75.5%) and in England & Wales (92/159 57.8%). The higher proportion of housed groups in Scotland could relate to typical weather conditions; there tends to be a longer period in the year during which animals can be kept outdoors in England & Wales.

In Scotland, few of the grazing sampled groups were mixed with other cattle (5/27, 18.5%) or with other livestock species (4/27, 14.8%). The other livestock species were mostly sheep. In England & Wales the picture was slightly different with a lower proportion of grazing sampled groups that shared grazing with other cattle (8/68, 11.8%), while more of the grazing sampled groups shared grazing with other livestock (15/68, 22.0%) – again, mostly with sheep.

A lower proportion of farms in the England & Wales survey (121/159, 76.1%) had moved cattle onto the farm in the past year, compared to the Scotland survey

(87/110, 79.1%). More of the farms in the Scotland survey held certified organic status (5/110, 4.5%) than in the England & Wales survey (3/159, 1.9%).

Season: The date of sampling was used to include the season of sampling in the risk factor analysis, as previous research has shown that the number of cattle with *E. coli* O157 in their faeces can be associated with different seasons and a number of management factors such as housing status may be associated with season (i.e. it is a potential confounder, or effect-modifier). The sampling year was divided into four seasons, to describe when farms were sampled. The seasons were made up of three-month blocks, with spring being March to May inclusive, and so on. In the England & Wales survey, difficulties with farmer recruitment led to fewer farms being sampled in the spring, and more being sampled in the autumn, than would have been proportional if sampling had been distributed evenly across the year.

Statistical analysis of risk factors: Simple statistical approaches were used to describe the information collected in the farm management questionnaires and so provide an overall picture of the types of farms sampled for this survey, as well as how the sampled animals were managed. The surveys were designed to estimate prevalence and provide strains of *E. coli* O157 that were representative of those currently circulating in the cattle population of Great Britain. They were not designed to collect sufficient quantities of data to formally investigate risk-factors for the presence of *E. coli* O157 at farm-level. This would have required a larger number of farms to be visited. We have therefore conducted only a basic statistical analysis of the factors that were associated with an increased chance of a farm in our surveys testing positive for *E. coli* O157. The results should be treated and interpreted with caution and in context.

Three outcomes of interest were investigated (Table 2.6). For both the Scotland and the England & Wales surveys, a primary list of potential risk factors (PRFs) for further investigation was drawn up. This was based on a brief literature review of previous risk factor studies and on biological factors considered to be relevant to *E. coli* O157. PRFs were sourced from both questionnaire and laboratory data. A screening process was used to select the PRFs which, based on a p-value of ≤ 0.20 for their association with the outcomes of interest would be retained for inclusion in a multivariable logistic regression model.

A forwards and backwards stepwise elimination approach was taken to arrive at a final multivariable model for each of the scenarios detailed in Table 2.6. This involves adding or removing variables one by one, to see what effect this has on the statistical model. For this stage of the process, a *p*-value of ≤ 0.05 was chosen as the significance threshold. Variables were retained in the final model if removing them or adding them back caused a statistically significant change in the deviance of the model.

Table 2.6. **The three different outcomes of interest that were investigated in the risk factor analysis.**

Analysis Level	Outcomes of interest		
Scotland	Farm classified as positive	Proportion of samples positive within positive farms	Presence of at least one super-shedder pat on positive farms
England & Wales			
Scotland + England & Wales; 'survey' included as an additional factor in the model			

Risk factor analysis results

The risk factors that were retained in the final model for each outcome along with an indication of the direction of their effect, are provided in Table 2.7, below.

Outcome i): farm classified as positive (Table 2.7)

In Scotland, being sampled in spring resulted in a statistically significantly lower chance of a farm testing positive for *E. coli* O157, compared to the autumn.

Farms that had bought livestock species other than cattle in the year preceding sampling were less likely to test positive in this survey, whilst those that had moved breeding female cattle onto the farm in that same time period were more likely to test positive. Both of these effects were statistically significant.

Finally, there was a statistically significantly increased chance of a farm testing positive for every additional bovine animal on the farm in the age group between 12 and 30 months.

In England & Wales, for every additional bovine animal present on the farm as a whole on the day of sampling, there was a significantly increased chance of that farm testing positive i.e. farms with more cattle were more likely to test positive.

When the statistical analysis was performed on both sets of survey results simultaneously, with a new variable "survey" to account for the influence of the individual surveys themselves, the major factor associated with farm positive status was the presence of greater total numbers of cattle on the farm.

Outcome ii): proportion of samples positive within positive farms (Table 2.7)

In both the Scotland and the England & Wales surveys, if the sampled group was housed at the time of sampling then this was associated with a statistically significantly increased proportion of the individual samples from that group testing

positive for *E. coli* O157, compared to if the sampled group was not housed at the time of sampling.

In the combined analysis, with the additional “survey” variable, the same effect of housing status remained significant overall. There was also a statistically significant effect of “survey”, with a decreased proportion of samples within positive farms testing positive in the England & Wales survey, over and above the effect of housing, compared to the Scotland survey.

Outcome iii): presence of at least one super-shedder pat in the sampled group on positive farms (Table 2.7)

In both of the surveys, there was a statistically significant association between the overall proportion of pats on positive farms that tested individually positive for *E. coli* O157 and the chance of there being at least one super-shedder pat on that farm.

When both surveys were modelled together the association of the presence of a super-shedder with the proportion of pats testing positive on the positive farms remained significant. There was no statistically significant effect of survey.

Table 2.7. **Final model for each outcome of interest with associated risk factors.** Arrows describe the direction of the effects based on the odds ratio (OR, at $p \leq 0.05$) in the final model. Survey is listed as a factor where Scotland data and England & Wales data were modelled simultaneously, even if not statistically significant, because it was required to be retained within those models.

Outcome	Final model risk factors		
	Scotland	England & Wales	Scotland + England & Wales
i) Farm classified as positive	Season: OR ↓ spring vs. autumn Bought livestock other than cattle in previous year: OR ↓ Yes vs. No Breeding female cattle moved onto the farm in previous year: OR ↑ Yes vs. No Number of cattle on farm between 12 and 30 months of age: OR ↑ with every additional animal	Total number of cattle on the farm: OR ↑ with every additional animal	Total number of cattle on the farm: OR ↑ with every additional animal Survey – no effect?
ii) Proportion of samples positive within positive farms	Housing status: OR ↑ if housed	Housing status: OR ↑ if housed	Housing status: OR ↑ if housed Survey: OR ↓ E&W vs. Scotland
iii) Presence of at least one moderate or high level super-shedder pat in a positive sample group	Proportion of samples positive: OR ↑ as prop_pos ↑	Proportion of samples positive: OR ↑ as prop_pos ↑	Proportion of samples positive: OR ↑ as prop pos ↑ Survey – no effect?

Interpretation of the outputs of the risk factor analysis

There were some differences in the farm characteristics and management systems between England & Wales and Scotland. These may contribute subtly to *E. coli* O157 status, although few statistically significant differences at farm level have been identified in the multivariable statistical models. They may also partially contribute to the finding that different factors appear to have an association with positive farm status on Scottish farms compared to farms in England & Wales.

Housing and season

In the UK, as well as elsewhere, housing status is often closely associated with season, as typical management approaches often lead to animals being housed when the weather is bad. This will vary from year to year and geographically. It is likely that farms in areas of England with better weather (i.e. milder and drier) may house cattle over a shorter time period, on average, than farms in Scotland. This could contribute to a possible explanation for the different effects of season and housing observed in the different models.

In the Scotland survey at farm-level, season is associated with the chance of a farm testing positive for *E. coli* O157. If a farm is positive, however, the proportion of samples within the sampled group that test positive is more likely to be influenced by whether or not cattle are housed.

The seasonal effect on *E. coli* O157 status at farm-level was not seen in England & Wales although the association with housing on the proportion of samples within the sampled group that test positive on positive farms was. The latter was consistent and remained in the combined model, although with an indication that there was a difference between the two surveys.

Housing of the sampled group of cattle is a factor that has been found to have an association with *E. coli* O157 status in other studies described in the scientific literature. It is biologically plausible that contact and thus transmission between animals is more likely when housed due to their close proximity, leading to more animals affected and shedding bacteria. Alternatively, the chance of survival of the bacteria in the pats may be improved by the environmental conditions, resulting in increased exposure.

This may be relevant to the discussion of how to mitigate the public health risk of *E. coli* O157 in cattle. Whilst the requirement to house cattle may be unavoidable, at least under current typical weather patterns and stocking densities, it may be possible to identify ways to intervene at the beginning of the housing season to reduce the burden of *E. coli* O157 at that point, in hopes that this could mitigate the risk to public health.

Presence of a super-shedder

The finding that the presence of at least one moderate or high super-shedder sample in a group is associated with a higher proportion of samples from the same group being positive is consistent in both of our surveys. It is plausible that the more positive samples there are within a group, the greater the chance that one will be in the super-shedder category. Equally, the presence of a super-shedder in the group is more likely to result in a greater number of animals within that same group shedding *E. coli* O157, as has been described in previous research[38].

This in turn lends supports to the idea that mitigation of super-shedding is likely to be an effective tool to combat the overall burden of *E. coli* O157 within cattle groups, and consequently the risk posed to humans from contact with cattle and their environments.

While these analyses provide some more insight into the dynamics of *E. coli* O157 in cattle and suggest certain factors that may be associated with farm positive status, the proportion of samples positive within positive groups, or presence of a super-shedder, it is important to acknowledge that the survey was not designed to formally investigate such risk factors. It may be that there was insufficient power to detect differences, due to the design of the study. Care should also be taken in the interpretation of the outputs as a cross-sectional survey design can only indicate hypotheses i.e. ideas for further investigation. It cannot lead to conclusions of causality.

The characteristics of the particular subtypes of *E. coli* O157 that are more prevalent on the Scottish versus English & Welsh farms are also relevant and described in sections 2.3.1-4.

2.5 Sequence analysis of human and cattle *E. coli* O157 isolates (DO 2.1.2-3 & 2.2-6)

2.5.1. Human isolates for sequencing

Scotland: Over the Scottish cattle sampling period (15/9/14 to 14/9/15), SERL received, or isolated from referred faeces, 220 *E. coli* O157 strains. Removing duplicate strains from the same patient, *stx* negative *E. coli* O157 and travel-associated strains, a study group of 163 STEC O157 strains were identified. The isolates were initially characterised using traditional methods (including phage typing, PCR and Multiple Locus Variable-number tandem repeat Analysis-MLVA).

The isolates represent the diversity of Scottish STEC causing clinical infection over a one-year period. These comprised 11 different phage types and three different profiles that were untypeable (RDNC) (one strain was subsequently identified as *stx* negative *E. coli* O157:H39). MLVA analysis of the 163 human strains identified 123 unique allelic profiles (for the purposes of this analysis, MLVA single locus variants were included as unique profiles).

England & Wales: 556 isolates were selected by PHE from England & Wales referrals for whole genome sequencing and analysis.

2.5.2. Bovine isolates for Illumina platform short-read sequencing

From the 2014/15 BECS survey, 112 isolates from cattle which represented at least one positive O157 isolate per PT per positive farm were selected for whole genome sequencing. A number of farms had multiple isolates with the same PT sequenced to address local diversity although this is best analysed by long-read sequencing (2.7). In addition, 96 isolates from the IPRAVE study were also selected for sequencing, focusing on both diversity of PT and farm origin. These were from 2002-3.

2.5.3. Short-read sequencing and analysis methods

DNA was extracted on the QIAsymphony platform (Qiagen) using the DSP DNA Mini Kit (Qiagen, Crawley, UK) following a pre-lysis step as recommended by the manufacturer. The purity and quantity of the DNA were measured with the NanoDrop ND-1000 (NanoDrop Products, Thermo Scientific) and the Qubit Fluorimeter 3.0 (Thermo Fisher Scientific) with the dsDNA Assay HS Kit. The DNA was then submitted to the PHE sequencing unit.

The genome sequences were processed through the routine PHE *E. coli* typing workflow which assigns species identification through a reference Kmer comparison (https://github.com/phe-bioinformatics/kmerid) and identifies the 7 gene Multi-Locus Sequence Type (MLST) through alignment of reads to alleles of the Achtmann[61] scheme using MOST (https://github.com/phe-bioinformatics/MOST). *In silico* serotyping was performed by aligning reads against the reference database of

Jonsen *et al*[62]. Virulence determinants were identified by mapping reads to the major pathotype delineating genes, namely *eae*, *ipaH, bfpA, aggR, aaiC, ltcA, sta1, stb* and *stx*. Shiga toxin subtyping was performed (https://github.com/flashton2003/stx_subtyping) and Single Nucleotide Polymorphisms (SNPs) identified by mapping reads to the reference genome Sakai using PHEnix (https://github.com/phe-bioinformatics/PHEnix) and SNP clustering performed using SnapperDB (https://github.com/phe-bioinformatics/snapperdb). Maximum likelihood phylogenies were constructed using RAxML[63].

2.5.4. Results of phylogenomic studies based on short-read sequencing (DO 2.2.2/6 & 7)

Previous analysis has shown that single linkage clustering based on pairwise genetic distance is an effective method of defining phylogenetic groups as it is inclusive of clonal expansion events. Using a SNP distance threshold of 250 (t250), the PHE database of approximately 4,800 isolates delineates into 98 broad clusters and generally agrees with the lineage assignments previously established for this organism.

STEC O157 worldwide is formed into three major lineages I, II and I/II and we can find all three lineages in the UK. Our data indicates that the current strains originate from a Stx2c ancestral strain some 175 years ago and this has subsequently been lysogenised by Stx2a carrying prophage into multiple strain backgrounds. These strains form the basis of highly pathogenic variants circulating both in the UK and worldwide today (OP2)[64].

In comparison to the previously described strain range, both cattle studies revealed a restricted set of diversities. The isolates from both the IPRAVE study and the Scottish BECS study were only associated with 7 t250 clusters with 4/7 overlapping the two study periods (Fig. 2.12), suggesting a stable population of *E. coli* O157 in Scottish cattle over the past decade. The proportion of isolates that fall into each cluster was broadly equivalent with t250:4 (lineage Ic-PT21/28) making up 61% of isolates in IPRAVE and 65% in the current BECS study. The second most prevalent cluster in both data sets was t250:2 (lineage IIc-PT8) accounting for 11% of isolates in the IPRAVE study and 16% in the current BECS study.

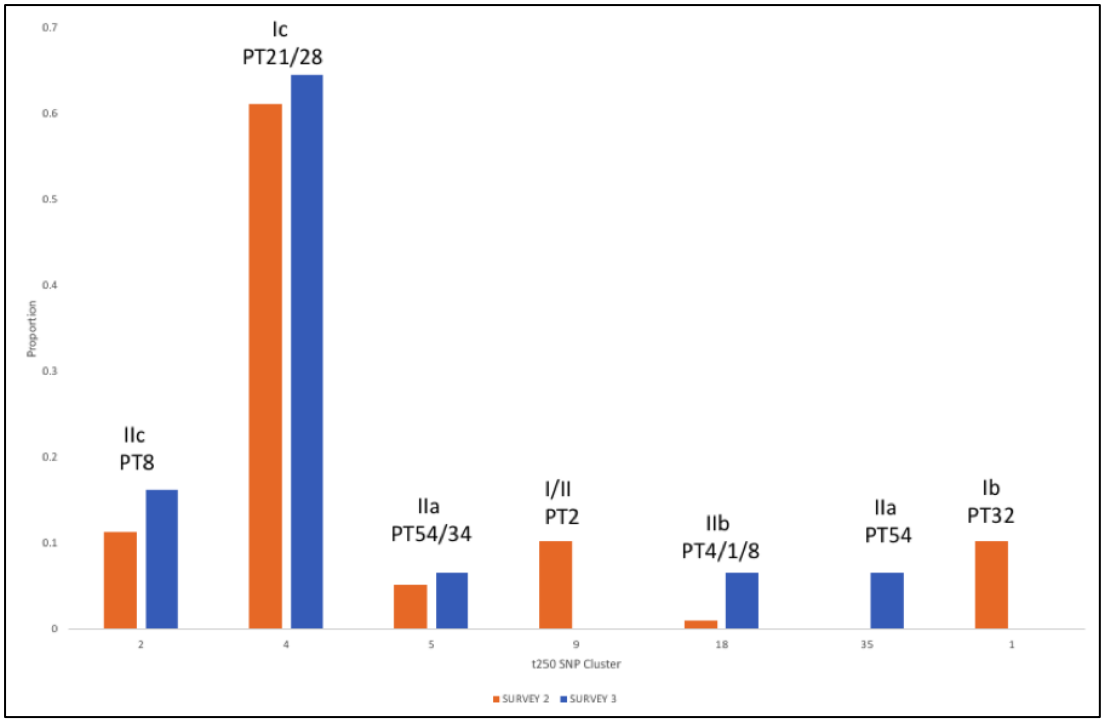


Figure 2.12 Histogram showing the proportion of t250 SNP clusters isolated in both the IPRAVE study (orange) and the BECS study (blue) in Scotland.

While the sampling was not designed to split analysis of England & Wales, it is potentially informative to compare the 2014/15 data across the three different countries involved to consider regional distribution of strain types. This reveals a much more heterogeneous distribution of diversity (Fig. 2.13). 7 t250 SNP clusters were isolated across the study sites, with t250:9 (I/II-PT2) and t250:107 (Ic-PT32) restricted to England and t250:35(IIa-PT54) only found in Scottish isolates. t250:2 (IIc-PT8) was found in all study sites in a consistent proportion in England and Scotland accounting for 14% and 16% of isolates respectively. t250:18 (IIb-PT4/1/8) accounted for over half (4/7) of the Welsh isolates and 6% and 13% of the Scottish and English isolates respectively. Two clusters displayed marked differences in proportions between Scotland and England. t250:4 (Ic-PT21/28) accounted for 65% of Scottish isolates but only 10% of English isolates, conversely t250:5 (IIa-PT54/34) was found in 50% of English isolates but only 6% of Scottish isolates.

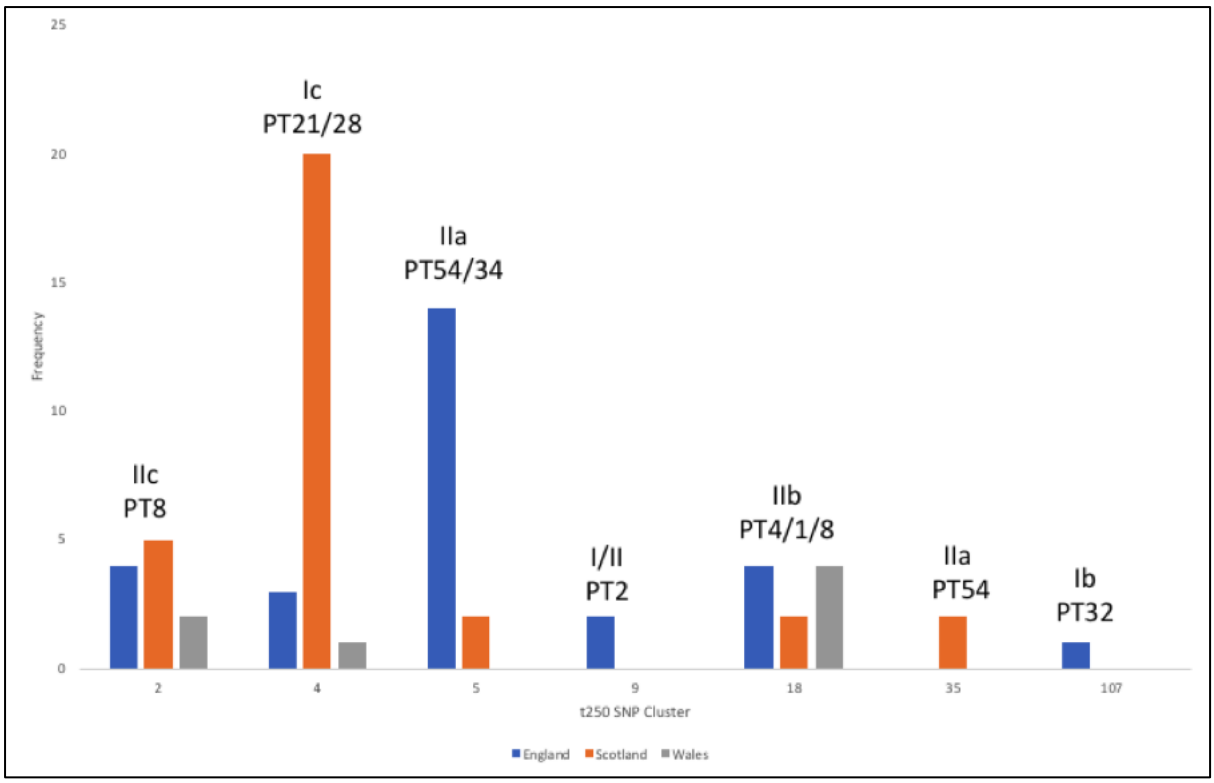


Figure 2.13. Histogram showing the number of t250 SNP clusters isolated in England, Scotland and Wales in the BECS study.

Fig. 2.14 shows the phylogenetic distribution of isolates from the IPRAVE study (Orange) and the BECS study (Blue). As demonstrated by the t250 SNP clustering, the isolates from the GB cattle surveys cluster across the three lineages but in restricted phylogenetic clades. The phylogenetic tree highlights regions where very similar isolates have been detected over both the IPRAVE and BECS studies but also regions of diversity sampled in the IPRAVE study but not in the current BECS study. This includes region A on the tree where the prototypical PT32 Stx2c harbouring strain 10671 clustered in the IPRAVE study. The BECS study also reveals regions of novel diversity not previously sampled including cluster B in t250:5(PT54) only found in English cattle. PT21/28 (t250:4) was found in high proportions (greater than 60%) in both studies in Scotland and from the phylogeny it can been seen the current diversity spans a large portion of the Ic lineage.

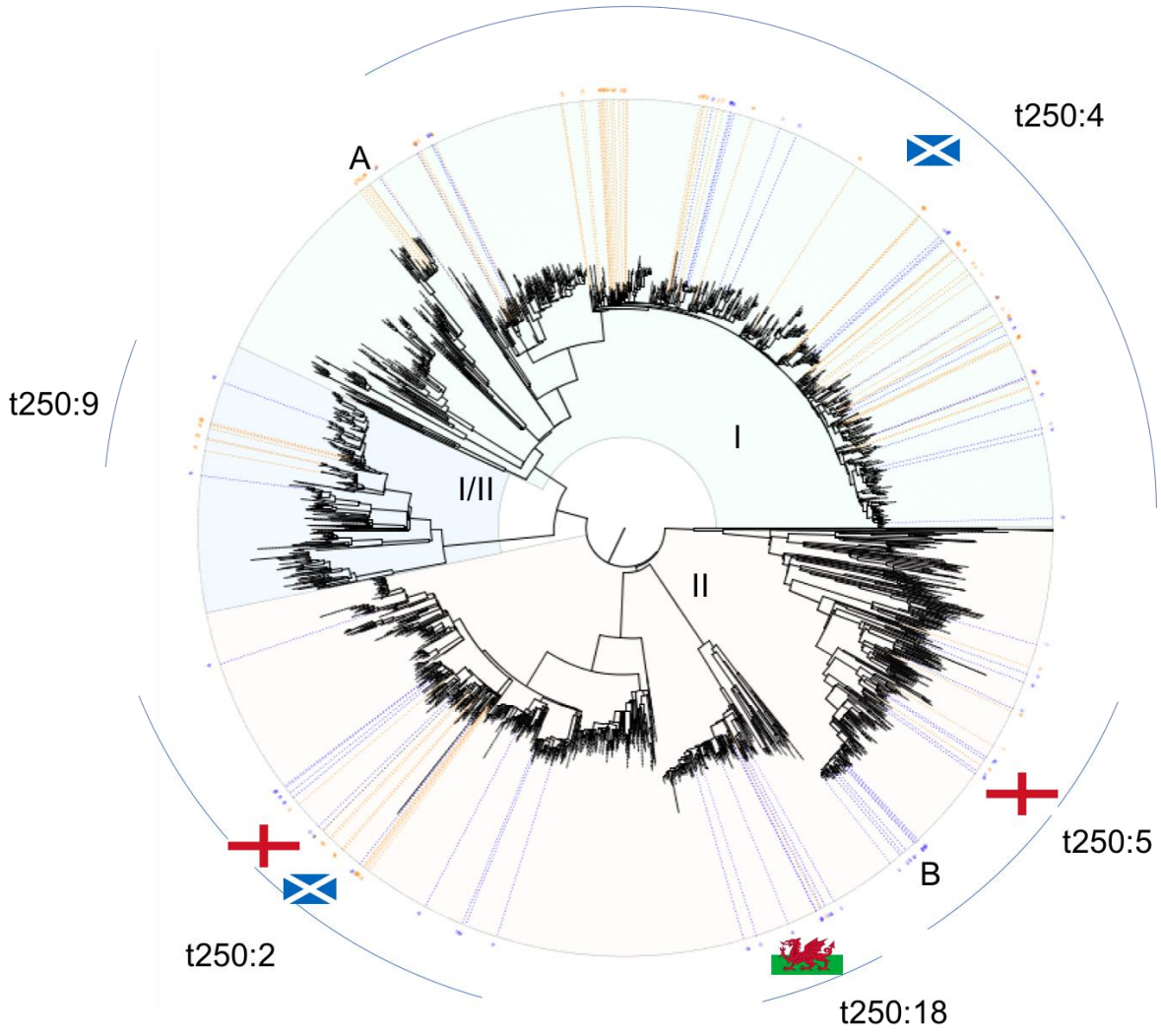


Figure 2.14. Maximum likelihood core genome SNP phylogeny of BECS IPRAVE study isolates against the global diversity. Blue BECS study isolates, Orange IPRAVE study isolates.

To investigate how the distribution of diversity corresponds to what was observed in human clinical cases, 163 isolates from Scottish routine surveillance and 556 clinical isolates from England & Wales were sequenced. Fig. 2.15 shows a core genome SNP phylogeny of the clustering of human clinical and cattle isolates, with seven outbreaks that occurred during the study period highlighted. Outbreaks 1 & 2 fall within lineage IIc (t250:2-PT8) and were both epidemiologically associated with the consumption of bagged salad of domestic origin[65]. These two outbreaks cluster within the diversity of surveyed cattle isolates sequenced during the study period. Outbreak 3 was also epidemiologically associated with bagged salad and emerged from the t250:18 (PT4/1/8) clade that contained cattle isolates from all three countries.

In Figure 2.14 it can be seen there are parts of the O157:H7 phylogeny that were associated predominantly with Scotland (t250:4–PT21/28) and conversely diversity associated more strongly with English cattle (t250:5-PT54). Although this may represent restricted diversity in the cattle population this diversity is sampled by human clinical cases in both England and Scotland. This is particularly apparent in the t250:4 cluster where human clinical cases from England are distributed throughout the diversity captured in the Scottish cattle and seen in the Scottish human clinical cases. Four outbreaks occurred in England & Wales involving PT21/28 strains (t250:4) in the study period (Fig. 2.15). This included OB4 associated with a butcher's shop and linked back to a cattle farm in the North East, OB5 associated with a private water supply in the North West and OB6 associated with sheep exposure at a petting farm[66]. In all three outbreaks the exposure could be linked to animal or environmental exposure within England. This highlights that despite t250:4 (PT21/28) not being isolated in high numbers in English and Welsh cattle the strain is undoubtedly present in these countries and raises the possibility of an un-sampled reservoir harbouring PT21/28 south of the border. OB7 was associated with the consumption of raw drinking milk[67] sourced from the South West of England and the surrounding clade contained the two PT21/28 strains isolated in English or Welsh farms. Furthermore, this part of the phylogeny contains a smaller proportion of clinical PT21/28 isolates from Scotland than the rest of the t250:4 cluster (Fig. 2.15). Analysis undertaken as part of the outbreak showed that the sporadic human clinical cases that clustered in this clade were also more likely to be resident in the South West of England providing further evidence of this strain circulating in English livestock in that region.

Another value of sequencing data is to be able to examine related strains and predict the likely common ancestor and its host. Bayesian evolutionary analysis by sampling trees (BEAST) was used to look at the recent ancestry of the PT21/28 phylogeny in the UK. It is striking that a pictorial summary of this analysis (Fig. 2.16) shows that all human isolates are likely to be founded from cattle strains, supporting the concept that strains circulating in the cattle population are the main reservoir of PT21/28 *E. coli* O157 human infection in the UK.

In conclusion, the data from this study provides a unique insight into the phylogenomics of *E. coli* O157 in Great Britain. We have shown that the genomic diversity in Scottish cattle has been largely maintained over the last decade. We have shown striking differences in composition of *E. coli* O157 north and south of the border, with Scotland having a consistently high proportion Stx2a harbouring t250:4 (PT21/28) while in England there was more variation and more isolates of the Stx2c t250:5 subcluster (PT54). This difference in strain composition in the cattle reservoir is, however, not reflected in the strain diversity found in the human population. Broadly, strains that are found in either English or Scottish cattle can be found in human clinical cases from both countries. This could represent a shared exposure

from both countries e.g. distributed food products or could represent an un-sampled reservoir that is maintaining the diversity.

In terms of public health, this survey represents a major advance in understanding which strains of O157 are likely to be from a domestic reservoir. This will support future outbreak investigations and hypothesis generation.

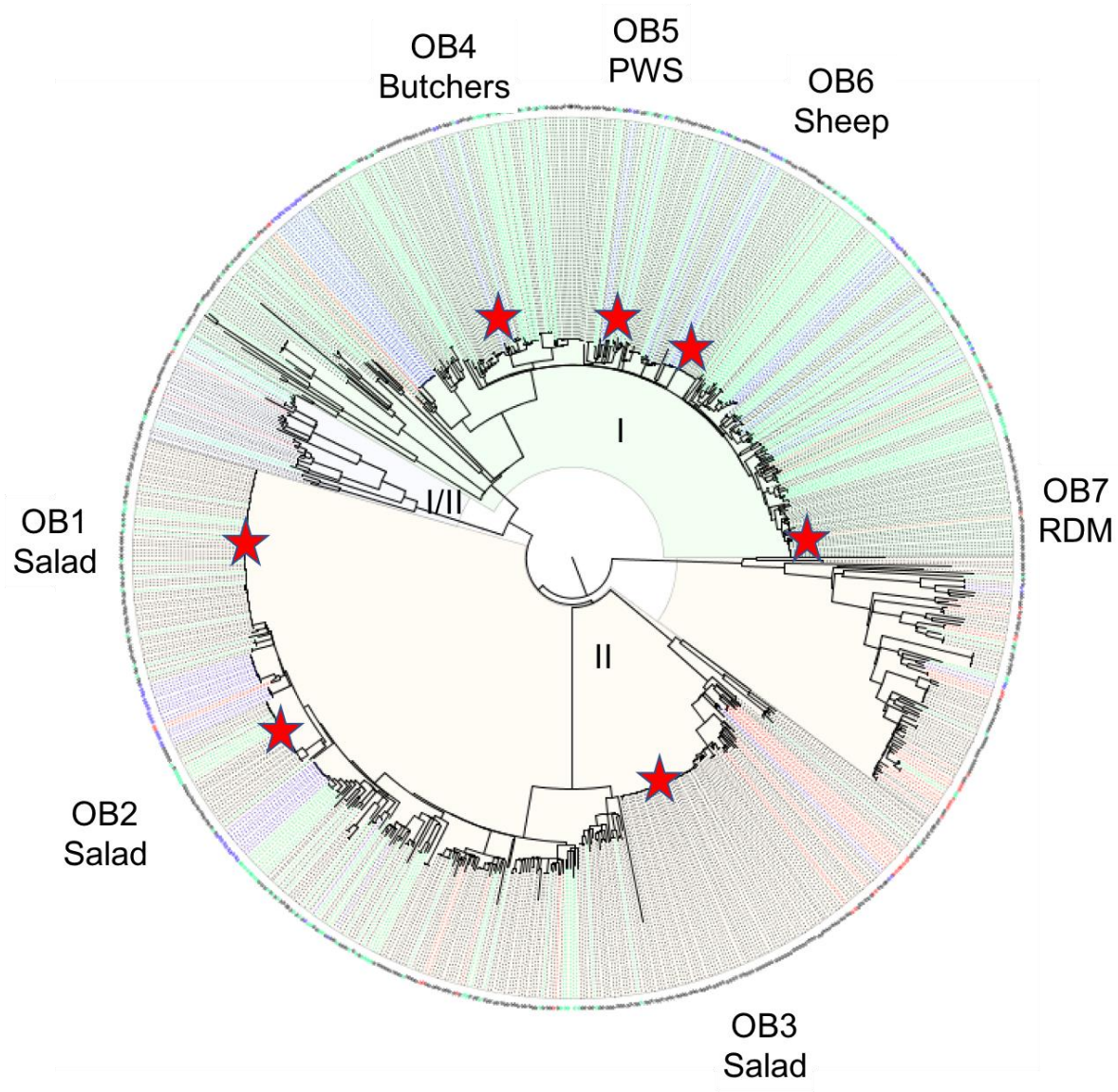


Figure 2.15. Maximum likelihood core genome SNP phylogeny of BECS study isolates against contemporaneous clinical isolates from England, Wales and Scotland. Black – England & Wales clinical isolates, Green – Scottish clinical isolates, Blue – Scottish Cattle, Red - English & Welsh Cattle. 7 outbreaks (OB) are marked on the phylogeny that occurred in the study period. OB1-3 associated with salad; OB4 a butcher's shop; OB5 a private water supply (PWS); OB6 with sheep and lamb, and OB7 with raw drinking milk (RDM).

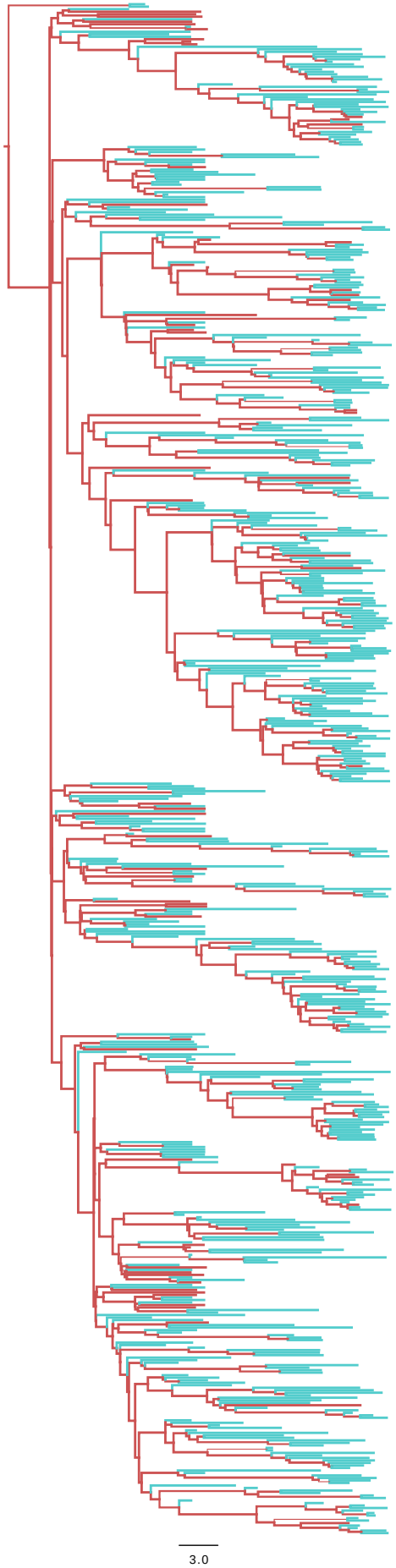


Figure 2.16. **Host prediction of recent ancestor analysis.** Bayesian evolutionary analysis by sampling trees (BEAST) was used to establish the predicted evolutionary ancestral state for the PT21/28 Lineage I cluster using 21 UK cattle and 602 human isolates. Branches indicating a cattle host (brown) are shown to lead to all human isolates (blue) at the tips of the tree. The majority of PT21/28 isolates associated with human infections have recent ancestors in cattle.

2.5.5. Host and zoonotic prediction based on machine learning approaches

Allied research on human and cattle sequences produced under this programme enabled us to investigate a machine learning approach (Support Vector Machine – SVM) to predict the likelihood that a cattle *E. coli* O157 isolate may be a threat to human health. As defined in the output publication (OP3)[68]:

> '*We demonstrate that only a small subset of bovine strains is likely to cause human disease, even within previously defined pathogenic lineages. The approach was tested across isolates from the UK and USA and verified with food and cattle isolates from outbreak investigations. This finding has important implications for targeting of control strategies in herds*.'

This work has recently been followed up by broader application of SVM to predict the host animal or environmental niche of *E. coli* and *Salmonella* Typhimurium isolates based on their genome sequence[69].

2.6. Whole genome sequencing (WGS) for routine diagnostics to enable SERL to transition to sequence-based diagnostics for *E. coli* O157 (DO 2.2.5-7)

A PhD studentship funded by the programme included work to deploy a Public Health England-designed Bioinformatics pipeline at the Scottish *E. coli* O157/STEC Reference Laboratory (SERL). Contemporaneously, the Scottish Reference Laboratory commissioners, National Services Division, sanctioned SERL to conduct a pilot study by validating the WGS process on strains of *E. coli* O157 and non-O157 STEC (n=150), using an Illumina MiSeq desktop sequencer based at Molecular Genetics, The Western General Hospital, Edinburgh.

The installation of the PHE Bioinformatics pipeline at SERL enables Scottish human isolates sequenced at SERL to be compared with those from England & Wales. The study was successful and standardisation of WGS analysis between the SERL and PHE was achieved, although data security issues prevent SERL and PHE both working 'live' to a cloud based reference database. Currently, SERL forward JSON files to PHE for further analysis and PHE then assign a UK SNP address to each isolate (SNP addresses only produced for *E. coli* O157 and other selected STEC) (APPENDIX E). A manuscript detailing this validation has now been published (OP4).

Whole-genome sequencing was introduced into routine use at SERL on 28th August 2017. Data is exchanged with PHE on a weekly basis in order to compare isolates. Sequence data is analysed using two different approaches at SERL, the PHE protocol, as described above, and also using BioNumerics (Applied Maths). This dual data analysis approach achieved UKAS accreditation in November 2017.

2.7. Sequencing: insights from long-read sequencing of *E. coli* O157 isolates using the Pacific Biosciences platform (relates to DO 2.2.1-9)

Prophage integration, excision, inter-phage recombination and insertion sequence (IS) element movement are considered a primary source of genomic variation in

E. coli O157. However, tracking such genomic evolutionary changes and understanding phage diversity has been problematic due to the inability of short-read sequencing to discern the large number of repetitive sequences shared by prophage or IS elements. In collaboration with Dr. James L. Bono (USDA, Nebraska) we have sequenced and fully assembled the genomes of > 50 *E. coli* O157 isolates using long-read Pac-Bio SMRT sequencing technology. Data generated has furthered our understanding of phage diversity and enabled specific clustering of Stx encoding phage[70] (OP5) and has aided public health investigations[71,72] (OP6 & 7). For example, work published during this study by researchers in Japan[73], has defined specific subgroups of Stx2a encoding prophages in-line with our published work[70], as expanded below. Ogura et al[73] go on to show that the different Stx2a prophage groups can be associated with different levels of toxin production. So being able to define the Stx2a encoding prophage subtype may have value in predicting the virulence of a strain and likely overlaps with prophage differences found by our machine learning approach[68].

The use of long-read sequencing during this funding period provided the opportunity to systematically catalogue and characterise the prophage population of *E. coli* O157 isolates. Nine strains were chosen for sequencing and used for analysis with an additional five publicly available *E. coli* O157 genomes. Isolates were selected from throughout the *E. coli* O157 phylogeny that were representative of the main lineages and different PTs. Based on gene content alone, we identified a total of 151 unique prophages present in the genomes of *E. coli* O157 isolates and showed chromosomal regional bias for their insertion. Specific analysis of Stx encoding prophage demonstrated that those encoding the highly toxigenic Stx2a variant were the most divergent group (5 clusters from 7 phage genomes), those encoding Stx1a were intermediate (3 clusters from 7 phage genomes) and Stx2c encoding prophage were the least divergent (2 clusters from 9 phage genomes)[70]. Stx-encoding phage sequences can provide information towards the origin of strains, even non-O157, that contain specific phage sequences[71]. We also found that the number of phage inserted IS elements significantly increased as phage genome size decreased and were frequently inserted in and disrupted genes required for phage excision and replication. This allowed us to propose an evolutionary model of phage entrapment and ‘fixing’ of phage whereby after initial entrapment of a phage by IS mediated disruption, non-essential gene groups (metabolism and transport, tRNAs, structural) are lost over evolutionary time while those that provide a selective advantage to the phage or host (recombination, replication) are retained[70].

On a broader scale, we have used long-read sequencing to investigate short-term genome evolution and genetic variation and concluded that large chromosomal rearrangements (LCR), rather than phage loss or acquisition, are also an important source of genomic variation between PT21/28 *E. coli* O157 strains. More long-read sequencing of other PTs needs to be carried out before we can know how common these types of rearrangements are among *E. coli* O157 isolates and other

genera/species. Fifty representative strains were chosen from throughout the PT21/28 core SNP phylogeny that included:

1. Strains isolated from a single farm
2. Strains with identical PFGE profiles (Profile C) but from different branches of the PT21/28 phylogeny, and
3. Isolates of the same strain collected following experimental colonisation of cattle.

A high level of synteny was observed across all PT21/28 strains, however significant genomic variation was evident as LCRs, typically inversions, centred at the terminus (Ter) of replication. Each LCR was bounded by homologous prophage regions that were maintained as inverted repeats in relatively symmetrical positions at either side of Ter. PFGE profile C strains had identical chromosomal arrangements indicating this was the originating arrangement for PT21/28 or that strains oscillate between arrangements and this is a commonly isolated conformation. Genome comparisons of isolates collected after experimental animal colonisation identified two variants that differed by 1.3 and 1.4 Mbp inversions, respectively, from the original inoculum strains suggesting such rearrangements are occurring within the animal host.

Preliminary transcriptional analysis of inverted variants identified numerous (>100) genes across two biological conditions that differed significantly from progenitor strains and work is underway to understand the biological significance of alternate genome conformations. We hypothesise that different conformations have the capacity to generate phenotypic variation and that key prophage have been maintained at specific genomic loci to provide recombinational regions that can generate these alternative conformations.

Finally, in collaboration with PHE, the project team has used long-read sequencing to differentiate *E. coli* O157 strains responsible for two outbreaks associated with the same restaurant in 2012[72]. The actual source of the bacteria was never defined but the investigation team considered it likely that the strains were imported in contaminated food with the larger outbreak possibly exacerbated by an infected or colonised food handler in the restaurant. The initial outbreak was associated STEC PT8 strain resulting in 4 cases while the second outbreak, eight weeks later, was associated with a STEC PT54 strain with >140 cases confirmed. Despite the change in PT, traditional MVLA typing suggested that the same STEC strain was responsible for both outbreaks and short-read sequencing demonstrated that the outbreak strains were closely related, differing by just three SNPs in the core genome. However, using Pac-Bio SMRT sequencing significant differences in phage content, plasmid content and genomic structure between the PT8 and PT54 outbreak strains were identified. Both outbreak strains harboured the pO157 virulence plasmid, however an additional incHII plasmid that conferred chloramphenicol and streptomycin antibiotic resistance had been acquired by PT54. Acquisition of this plasmid was also responsible for the PT8/PT54 phage type switch. PT54 had also acquired an additional prophage that was not identified in any PT8 outbreak isolate

and encoded 30 unique genes. These genomic differences likely contributed to the increased pathogenicity of PT54 which was also found to be more acid resistant under certain physiologically relevant conditions.

Long read sequencing can therefore provide information on the accessory genome. Specifically, accurate structures of prophages and plasmids, as well as the overall organisation of the genome. *E. coli* O157 isolates, despite belonging to a clonal group, are still diverse and constantly evolving, especially through loss and acquisition of prophages and plasmids as well as genome rearrangements. These will contribute to variation in the colonisation/excretion potential from cattle and virulence of an isolate in a human host. Therefore, this variation needs to be monitored and understood in terms of its relationship to the zoonotic threat of the isolate.

2.8. Conclusions from Objective 1

This extensive work has established that while the overall prevalence of *E. coli* O157 on farms is equivalent in England & Wales vs Scotland at both herd and individual animal level, there is a clear inconsistency with human infection data for which the risk of acquiring an *E. coli* O157 infection is greater in Scotland with cases per 100,000 being approximately double those in England & Wales. An analysis of the strains that are present on farms shows that Scotland has a significantly reduced diversity of *E. coli* O157 types with one specific type (PT21/28) predominating, especially in northern Scotland. This type generally encodes Stx2a and Stx2c and we know that Stx2a is the most important risk factor for serious human infection. As such, we propose that the higher human rate in Scotland is a result of more local exposure to Stx2a+ *E. coli* subtypes in Scotland, led by PT21/28. Sequencing enables further definition of the *E. coli* isolates that are present in animals and humans which highlights the specific differences between strains that are more of a threat to human health.

3. Objective 2: Excretion dynamics and transmission frequencies of wild type *E. coli* O157 strains under controlled experimental conditions.

3.1 SUMMARY

A series of controlled challenge studies were performed to define the transmission dynamics of representative PT21/28 and PT32 *E. coli* O157 strains, as well as determining the role of Stx subtype 2a (Stx2a) on bacterial shedding levels from colonised calves and calf-to-calf transmission. Immune response measurements were performed on colonised calves. Finally, the effect of Stx2a on bacterial survival through the rumen was evaluated to determine whether Stx2a was involved in killing ruminal protozoa which are known to predate these bacteria. These studies conclude the following:

- Stx2a was associated with enhanced cattle-to-cattle transmission of *E. coli* O157.
- Both Stx2a positive and negative PT21/28 strains were shed from colonised cattle at higher levels and were more transmissible between cattle than the PT32 strain. This suggests that factors other than Stx2a within the PT21/28 strain also contribute to higher shedding levels and cattle-to-cattle transmissions.
- Increased transmission rates associated with Stx2a were not associated with enhanced survival of *E. coli* O157 within the rumen or suppression of local antibody responses at the rectal site of colonisation.
- Enhanced transmission rates associated with Stx2a are likely to account for the successful emergence of Stx2a+ *E. coli* O157 isolates in cattle in the last 40 years.

3.2. Experimental design for excretion and transmission studies

Characteristics of *E. coli* O157 strains used in experimental cattle studies

A summary of the three *E. coli* O157 strains used in the animal studies is shown in Table 3.1. Representative PT32 and PT21/28 strains were selected for initial studies for the following reasons: The PT32 strain 10671 and PT21/28 strain 9000 were isolated from bovine faecal pats as part of the IPRAVE national survey.

Strain 9000 was selected as it was associated with a high single pat count (6.9 x 10^5 cfu/g) while strain 10671 was only detected by enrichment in a faecal pat (<100 cfu/g). Strain 9000 possessed both a Stx2a and Stx2c encoding prophages whereas 10671 possessed a Stx2c-encoding prophage only.

Prior to the start of this project, both strains were sequenced using short-read sequencing technology. Alignment of whole genome sequences for strains PT21/28 strain 9000 and PT32 strain 10671 indicated that the primary difference between the two strains is the integrated Stx2a encoding prophage in strain 9000; other differences were also detected including 2829 SNP differences between the two strains and a total of 315 and 188 genes unique to strains 9000 and 10671, respectively, and other prophage differences (Fig. 3.1).

Table 3.1. Details of *E. coli* O157 strains used for experimental calf studies.

Strain	Functional *stx* genes	Phage Type	Origin	Trial
9000	*stx2c*	PT21/28	Cattle faeces (pat count 6.9 x 10^5 cfu/g)	1
10671	*stx2c*	PT32	Cattle faeces (pat count <100 cfu/g)	2
9000R	*stx2a, stx2c*	PT21/28	Derived *in vitro* from strain 9000	3

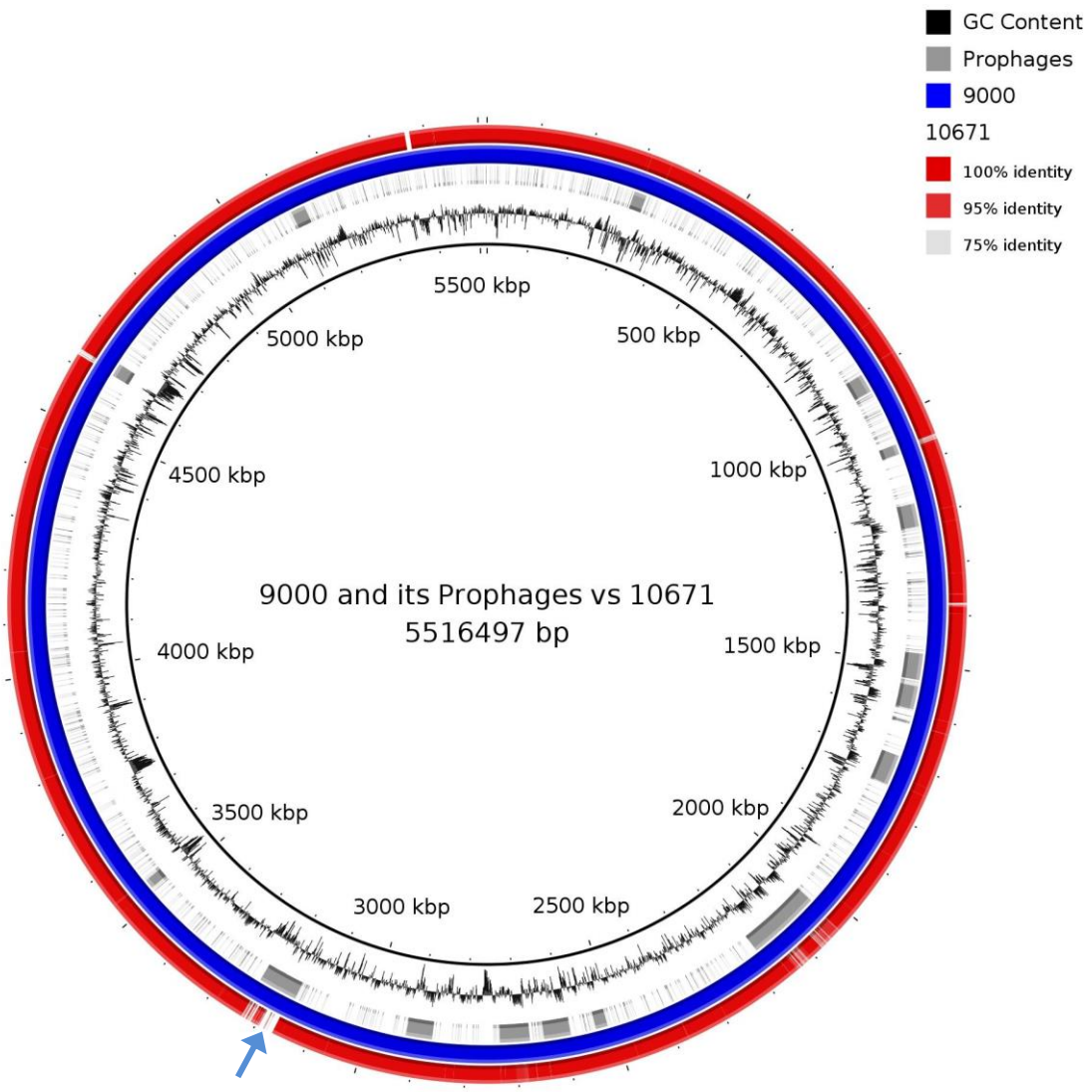


Figure 3.1. Genomic comparison of strain 9000 (PT21/28) and strain 10671 (PT32) used in the animal studies. The whole genome sequence of PT21/28 strain 9000 (blue) was used as a reference against which PT32 strain 10671 (red) was compared for gene presence/absence using a BRIGG plot. The annotated prophage from strain 9000 (grey) are also shown, including the Stx2a encoding prophage centred at 3,200 kbp shown by the blue arrow.

During the course of this project, both strain 9000 and 10671 were submitted for long-read sequencing. This analysis identified an ISEc8 insertion sequence within the *stx2a* subunit A gene of strain 9000 which was not apparent using short-read sequencing. Therefore this strain was only capable of producing functional Stx2c toxin, and not Stx2a toxin as was previously thought. As testing of a Stx2a strain was a key element of this objective, we generated an isogenic strain of 9000 (strain 9000R) which was capable of producing functional Stx2a and Stx2c toxins by precisely removing the ISEc8 from the *stx2a* gene by allelic replacement.

Sequencing was performed to confirm removal of ISEc8 and a sequence alignment of the Stx2a prophage from strains 9000 and 9000R is shown in Fig. 3.2.

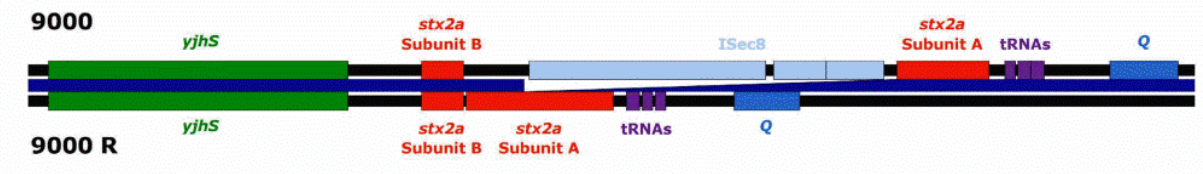


Figure 3.2. Stx2a prophage sequences indicating the presence of ISEc8 within the *stx2a* A subunit gene of strain 9000. Genomes (black lines), named genes (coloured blocks) and regions of homology (indigo lines) are shown.

Prior to testing in animals, Stx2 expression for strains 9000, 9000R and 10671 was determined *in vitro* following phage-induction. Production of Stx2, as determined by a pan-Stx2 ELISA which detects both Stx2a and Stx2c, was 2-fold higher in strain 10671 than strain 9000 (Fig. 3.3). Upon deletion of ISEc8 from the Stx2a Subunit A gene in strain 9000R, pan-Stx2 toxin production in strain 9000R was over 10-fold higher than strains 10671 or 9000. This confirmed that removal of the ISEc8 now resulted in production of Stx2a in addition to Stx2c, and that Stx2a is expressed at higher levels than Stx2c (Fig. 3.3).

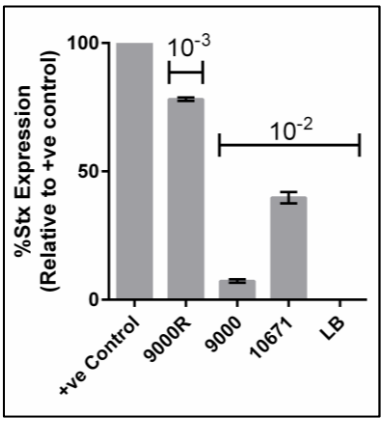


Figure 3.3. Stx2 production by strains used in animal studies. Total Stx2 toxin produced by PT21/28 strain 9000, 9000R and PT32 strain 10671 was measured by ELISA assay after 24 hrs of phage induction with Mitomycin C. Values are expressed relative to a positive control provided in the ELISA assay kit. The mean ± SEM from four biological replicates (n = 4) are shown for each strain. 10^{-3} and 10^{-2} indicate the dilution factor for the samples used for the Stx2 toxin ELISA assay (RIDASCREEN® Verotoxin ELISA kit (Biopharm). Positive control = inactivated verotoxin culture provided by the ELISA assay kit. LB = Lysogeny Broth without bacteria.

To determine whether higher Stx2 expression is a common feature of PT21/28 strains encoding Stx2a *vs*. Stx2c, total Stx2 production was determined for ten PT21/28 strains encoding Stx2a (with or without Stx2c) and ten PT21/28 strains encoding Stx2c alone. This indicated that PT21/28 strains encoding Stx2a generally produced higher levels of Stx2 than those encoding Stx2c alone (Fig. 3.4).

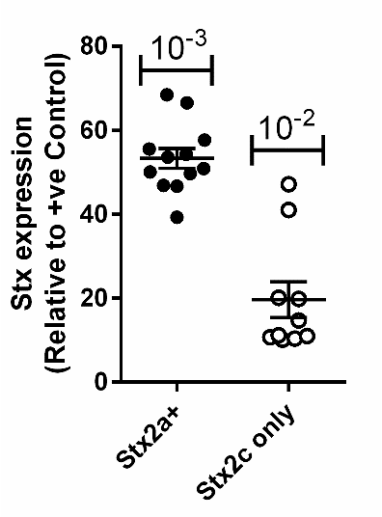


Figure 3.4. Total Stx expression is enhanced in PT21/28 strains encoding Stx2a with or without Stx2c compared to those encoding Stx2c alone. Following induction by Mitomycin C, total Stx production higher in PT21/28 strains encoding Stx2a with or without Stx2c (Stx2a+, n=10) *vs.* those encoding Stx2c alone (Stx2c only, n=10). 10^{-3} and 10^{-2} = dilution factor for the Stx2 ELISA assay.

Animal study design

Excretion and transmission dynamics of the three *E. coli* O157 strains detailed in Table 3.1 which differed in their Phage Type (PT) and/or Stx2 repertoire were determined under controlled experimental conditions within the Moredun Research Institute High Security Unit (HSU), following approval by the Moredun Experiments and Ethical Review committee. For all studies, conventionally reared male Holstein-Friesian calves were used with an average age of 12 ± 2 weeks at the start of the experiment. Calves were screened weekly and confirmed negative by both immunomagnetic separation (IMS) for *E. coli* O157 (using anti-*E. coli* O157 Dynabeads; Thermo Fisher) and quantitative PCR for detection of *E. coli* O157:H7 *rfb* gene and *stx1* and *stx2* for four weeks prior to the start of each trial.

The experimental study design is shown in Fig. 3.5. At Day 0 all calves housed in room C1 of the MRI HSU were challenged by orogastric intubation with ~10^9 cfu of nalidixic acid resistant (Nal^R) derivatives of the *E. coli* O157 strains (Trial 1: Strain 9000, Trial 2: Strain 10671, Trial 3: Strain 9000R). Five days after oral challenge, two high shedding calves (> 10^4 cfu/g faeces) were moved, one into room C2 and one into C3, each housing *E. coli* O157 naïve sentinel animals.

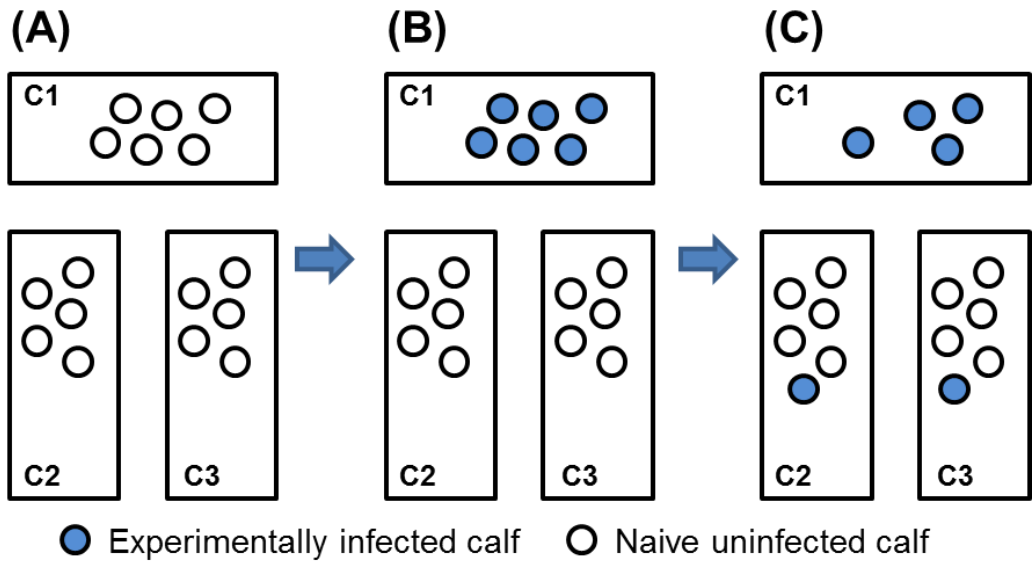


Figure 3.5. Experimental design for *E. coli* O157 transmission studies. *E. coli* O157 naïve calves were housed in three separate rooms (C1, C2 and C3) within the MRI HSU as detailed in (A). All calves in C1 were challenged by orogastric intubation with ~10^9 *E. coli* O157 (B). Five days later, a calf shedding high levels of *E. coli* O157 (>10^4 cfu/g faeces) was moved into C2 and C3 (C). Faecal bacterial shedding from all calves, as well as levels of bacteria swabbed from the floors of each pen, were monitored for a further 20-day period.

Faecal shedding was monitored in all challenged and sentinel animals over a 25-day and 20-day period respectively, as summarised in Table 3.2. For enumeration of bacteria, a direct plating technique was employed in which serial 10-fold dilutions of faecal samples on cefixime-tellurite sorbitol MacConkey (CT-SMAC) agar supplemented with nalidixic acid. Five to 10 colonies from each plate were confirmed O157 positive using an *E. coli* O157 Latex Test kit (Thermo Fisher). Where no colonies were observed, broth enrichment was performed by incubating faecal samples overnight at 37 °C in Tryptone Soya Broth (TSB; Oxoid). Enriched samples were then plated onto CT-SMAC plates supplemented with nalidixic acid and any bacterial growth tested for the presence of O157 by latex agglutination. Faeces negative by direct plating but positive after enrichment were assigned an arbitrary value of 10 cfu/g. Throughout the trial environmental samples from the floors of the pens were collected using absorbent swabs attached to overshoes. These environmental samples, which consisted of bedding and faecal material, were analysed as described for the faecal samples and the data expressed as counts per gram of environmental material.

Table 3.2. Summary of methods used to determine *E. coli* O157 shedding in cattle

Test	Samples analysed	Measurement obtained
Direct plating	Serial dilutions of faecal or environmental samples	Colony forming units per gram of faeces (cfu/g)
Broth Enrichment	Faecal or environmental sample cultured in broth to enrich for *E. coli* bacteria	*E. coli* O157 positive (>0 and <33 cfu/g) or negative. Positive results assigned as 10 cfu/g.

3.2. Experimental design for excretion and transmission studies

Characteristics of *E. coli* O157 strains used in experimental cattle studies

A summary of the three *E. coli* O157 strains used in the animal studies is shown in Table 3.1. Representative PT32 and PT21/28 strains were selected for initial studies for the following reasons: The PT32 strain 10671 and PT21/28 strain 9000 were isolated from bovine faecal pats as part of the IPRAVE national survey.

Strain 9000 was selected as it was associated with a high single pat count (6.9×10^5 cfu/g) while strain 10671 was only detected by enrichment in a faecal pat (<100 cfu/g). Strain 9000 possessed both a Stx2a and Stx2c encoding prophages whereas 10671 possessed a Stx2c-encoding prophage only.

Prior to the start of this project, both strains were sequenced using short-read sequencing technology. Alignment of whole genome sequences for strains PT21/28 strain 9000 and PT32 strain 10671 indicated that the primary difference between the two strains is the integrated Stx2a encoding prophage in strain 9000; other differences were also detected including 2829 SNP differences between the two strains and a total of 315 and 188 genes unique to strains 9000 and 10671, respectively, and other prophage differences (Fig. 3.1).

3.3 Results for Objective 2

3.3.1. Excretion dynamics of PT21/28 and PT32 strains in orally challenged calves (DO 1.1.1-3 & 1.1.6-7).

Mean bacterial shedding from calves orally challenged with strains 10671 (PT32, Stx2c+, n = 6), 9000 (PT21/28; Stx2c+, n = 4) and 9000R (PT21/28; Stx2a+ Stx2c+, n = 7) is shown in Fig. 3.6A. A Poisson Generalised Linear Mixed Model (GLMM) was fitted to statistically determine strain specific differences in mean cfu/g counts over time, as this model took into account time-dependent changes in bacterial shedding between the three strains.

A statistically significant ($p < 0.001$) interaction effect between strain and time was observed with the mean counts of strain PT32 10671 predicted to diverge over time

from strains 9000 and 9000R (Fig. 3.6B). Pairwise testing of the differences in mean cfu/g between the strains from the GLMM estimates further supported this conclusion, indicating significantly lower daily mean bacterial shedding from calves challenged with strain 10671 compared to either strain 9000 or 9000R ($p = 0.012$ and $p = 0.018$, respectively), but no statistically significant differences in shedding between strain 9000 and 9000R ($p = 0.454$). Using Area Under the Curve (AUC) analysis to estimate total bacteria excreted over the duration of each trial and fitting a negative binomial Generalised Linear Model (GLM) for strain comparison (to account for the distribution of AUC data not being normally distributed, but consisting of many low and a few high AUC values) no statistically significant differences in mean bacteria excreted were observed between the three strains ($p = 0.108$), although this is likely to be due to a lack of statistical power (Fig. 3.7). Thus, in our oral challenge model, both PT21/28 strains, 9000 and 9000R were excreted at significantly higher mean levels than PT32 strain 10671 over time.

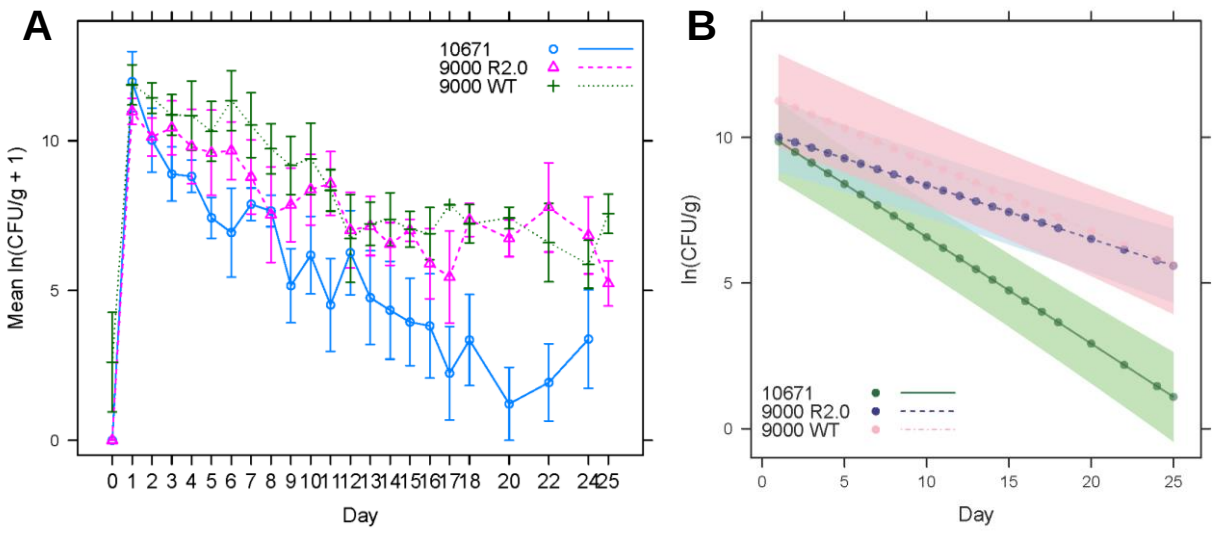


Figure 3.6. Analysis of mean daily shedding from experimentally challenged calves. Data is presented from calves orally challenged with *E. coli* O157 strains 10671, 9000 and 9000R. (A) The mean ± SEM daily cfu/g faeces (in log scale) from animals challenged with strains 10671 (n = 4) (blue), 9000R (n = 7) (pink) and 9000 (n = 6) (green) are shown. Daily counts for individual animals were calculated from three sets of dilution plates and used to calculate the daily mean for each strain. (B) A Poisson GLMM was used to determine strain-specific differences in mean cfu/g over time. Predicted mean cfu/g values ± 95 % confidence interval (CI) bands (in log-link scale) are shown for 10671 (red), 9000R (green) and 9000 (blue).

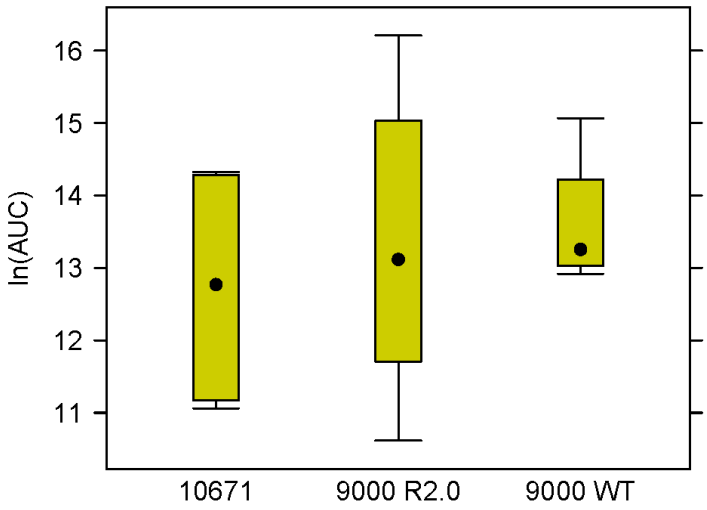


Figure 3.7. Analysis of total shedding from experimentally challenged calves. Total shedding (estimated by AUC) of each strain was calculated and compared using a negative binomial GLM. Predicted mean AUC ± 95 % CI (in log-link scale) are shown.

3.3.2. Transmission of PT21/28 and PT32 strains between cattle (DO 1.2.1-4, 1.3.1/3)

Transmission events between experimentally challenged and sentinel animals were monitored by enumerating cfu/g faeces for all sentinel animals in rooms C2 (n = 5) and C3 (n = 5). The mean daily cfu/g from animals colonised by strains 10671, 9000 and 9000R are plotted in Fig. 3.8. No transmission of PT32 strain 10671 from challenged to sentinel animals was detected, even by sample enrichment of faecal samples for *E. coli* O157, over the trial duration and this strain was therefore excluded from subsequent statistical analysis. In contrast both PT21/28 strains, 9000 and 9000R were successfully transmitted from challenged animals to 9/10 and 10/10 sentinel animals, respectively. Furthermore, 6/10 sentinel animals colonised with 9000R subsequently became super-shedders many of which excreted > 1 x 10^3 cfu/g for several days (Fig. 3.8). Only one super-shedding event was recorded for sentinels infected with strain 9000 for which detection was primarily sporadic and generally required sample enrichment (Fig. 3.8).

As above, statistical modelling was used to determine strain specific differences in mean cfu/g counts over time and total excretion from sentinel animals for strains 9000 and 9000R. Strain 9000R was predicted to produce statistically significantly higher mean cfu/g counts over time in the sentinel animals compared to strain 9000 ($p < 0.001$, Fig. 3.8B). A statistically significant strain × time interaction effect ($p < 0.001$) was identified, indicating that the difference in shedding from sentinels between the two strains was estimated to increase over time. The mean total excretion (AUC) was also statistically significantly ($p < 0.001$) higher from sentinel animals colonised by strain 9000R than those colonised by strain 9000 as the mean

total cfu/g for strain 9000R was estimated to be ~50-times greater than 9000 (Fig. 3.9). PT21/28 strains 9000 and 9000R are isogenic, as confirmed by Pac-Bio WGS, with exception of ISEc8 which had been removed from the *stx*2a A subunit gene in strain 9000R. Removal of ISEc8 in strain 9000R significantly increased Stx2a toxin production and toxicity (Fig. 3.4). **We therefore attribute the significantly higher cfu/g over time and total shedding from sentinel animals colonised by strain 9000R in our transmission model to the expression and activity of Stx2a toxin**. The lack of any difference in bacterial shedding between calves orally challenged with a high dose of strain 9000 and 9000R most likely due to the high challenge dose (10^9 cfu) used, which would compensate for reduced colonisation efficiency of the 9000 strain. Interestingly, in all studies, mean environmental levels of *E. coli* O157 within each pen largely mirrored the mean bacterial shedding from calves within the pens (Fig. 3.10).

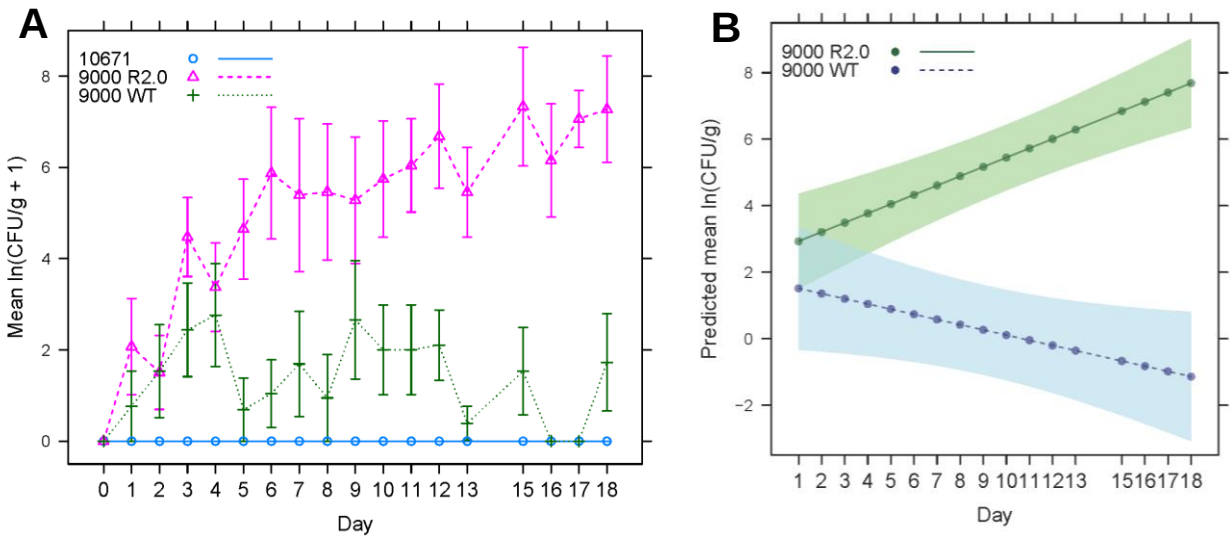


Figure 3.8. Analysis of mean daily shedding from sentinel calves. Data is presented from sentinel calves exposed to calves shedding *E. coli* O157 strains 10671, 9000 and 9000R. (A) Mean daily ± SEM cfu/g faeces (in log scale), (B) GLMM predicted mean daily cfu/g ± 95 % CI bands (in log scale).

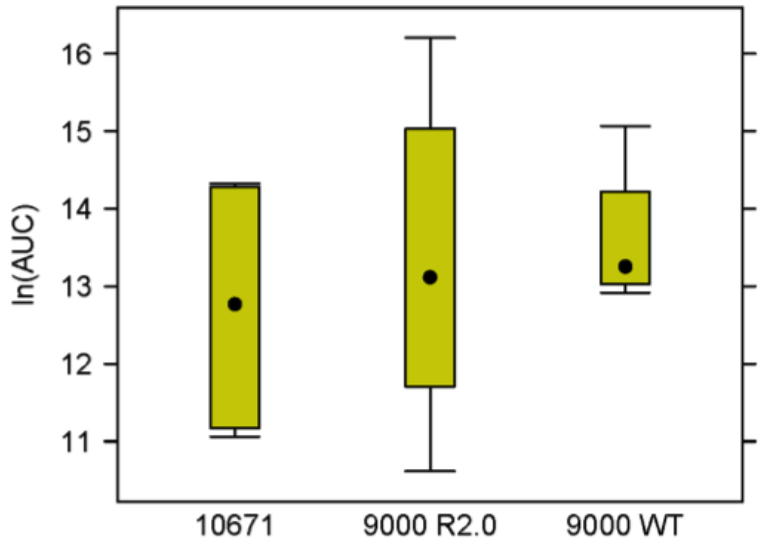


Figure 3.9. Analysis of total shedding from sentinel calves. Total shedding (estimated by AUC) of each strain was calculated and compared using a negative binomial GLM. Predicted mean AUC ± 95 % CI (in log-link scale) for sentinel animals are shown.

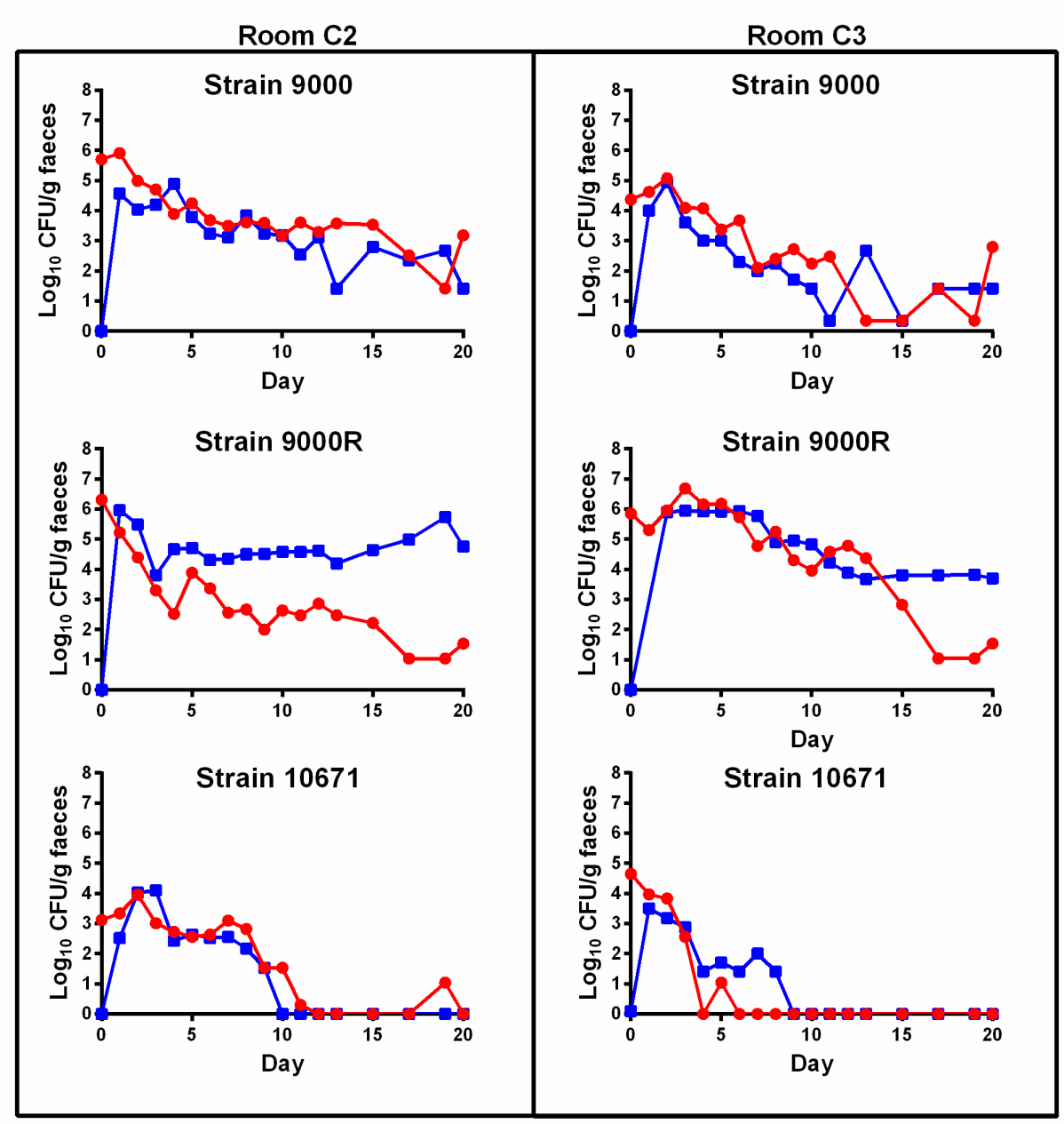


Figure 3.10. The relationship between shedding of *E. coli* O157 from experimentally infected calves and levels of *E. coli* O157 in the environment. Bacterial shedding was monitored from experimentally challenged animals moved into rooms C2 and C3 for three strains of *E. coli* O157 (9000, 9000R and 10671) and is indicated in red. Environmental levels within each room were also recorded and are indicated in blue. Data points represent the cfu/g faeces (for calves) or per gram of environmental material (for environmental levels) from three replicate plate counts.

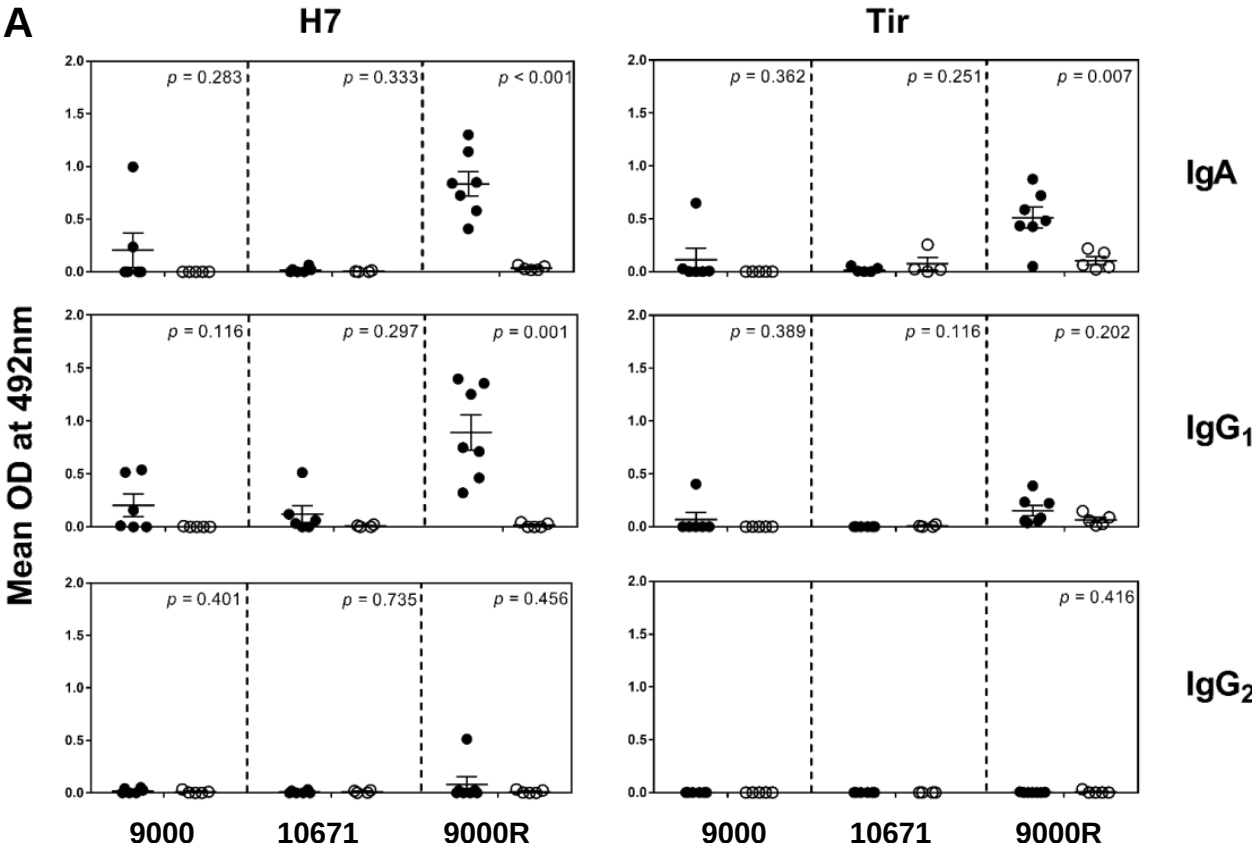
3.3.3 Immune response to E. coli O157 in challenged calves (DO 1.1.4/5 & 1.3.2)

In our previous studies it was clear that Stx2a was capable of enhancing calf-to-calf transmission. As Stx2 has been previously shown to suppress adaptive immune responses in cattle[19,20,74], one possible explanation for this enhanced transmission was that Stx2a was able to suppress *E. coli* O157-specific adaptive immune responses within the intestinal mucosa, thus allowing more efficient colonisation of the bovine intestinal tract. We therefore characterised circulating and rectal cellular and humoural *E. coli* O157-specific immune responses in calves orally challenged with all three strains, with the hypothesis that adaptive immune responses to strain 9000R, which had a functional *stx*2a and produced the highest levels of Stx2 in vitro, would be lower than those induced by strains 9000 or 10671. Circulating lymphocytes were analysed as these will contain T-cells which have been activated within the gut and re-circulate via the lymphatics and blood before exerting their effector functions within the intestinal tract and would allow the dynamics of the T-cell response during challenge to be characterised. Rectal lymph node cells were also analysed at post-mortem to characterise T-cell responses at the site of *E. coli* O157 colonisation.

***E. coli* O157-specific cellular immune responses:** Peripheral blood mononuclear cells (PBMCs), which consist of ~95% lymphocytes, were isolated from weekly blood samples and were re-stimulated with type-three secretion system protein preparations (T3SP) generated from *E. coli* O157. Antigen-specific lymphoproliferative responses and interferon-gamma (IFN-γ) production were assessed using a lymphocyte proliferation assay (LSA) and ELISpot assay, respectively. Lymphoproliferative responses were assessed in rectal lymph node cells collected at post-mortem. No statistically significant differences in either lymphoproliferative responses or numbers of antigen-specific IFN-γ producing cells were observed in challenged vs. control calves for any of the three challenge strains in either the blood or within rectal lymph nodes, indicating that *E. coli* O157 challenge with all three strains tested induced a limited *E. coli* O157-specific cellular immune response (data not shown).

***E. coli* O157-specific antibody responses:** Antibody responses to the following four *E. coli* O157 antigens: H7, EspA, Intimin and Tir, were determined by ELISA in weekly serum samples and in antibody secreting cell (ASC) probes generated from rectal lymph node cells collected at post-mortem. Challenge with all three strains induced a significant increase in serum H7-specific IgA ($p < 0.01$). An increase in serum levels of H7-specific IgG_1 was also seen in calves challenged with strain 9000R only ($p = 0.029$). A small increase in Tir and EspA-specific IgG_1 was also observed in calves challenged with strain 10671 ($p = 0.045$ & 0.023 for Tir and EspA, respectively) but not the 9000 or 9000R strains.

Antibody responses in the rectal lymph node which drains the site of colonisation of *E. coli* O157 are shown in Fig. 3.11. Challenge with strain 9000R (PT21/28 *stx*2a+ *stx*2c+) resulted in significantly higher levels of H7, Tir and EspA specific IgA ($p < 0.001$, 0.007, 0.025, respectively), and H7-specific IgG_1 ($p = 0.001$) responses in the rectal lymph node. No significant changes in rectal antibody responses to any of the *E. coli* O157 antigens tested were detected following challenge with either strains 10671 (PT32 *stx*2c+) or 9000 (PT21/28 *stx*2c+). This indicates that local *E. coli* O157-specific antibody responses were greatest following challenge with strain 9000R and suggests that the enhanced transmission of the PT21/28 strains associated with Stx2a is not due to suppression of *E. coli* O157-specific immune responses.



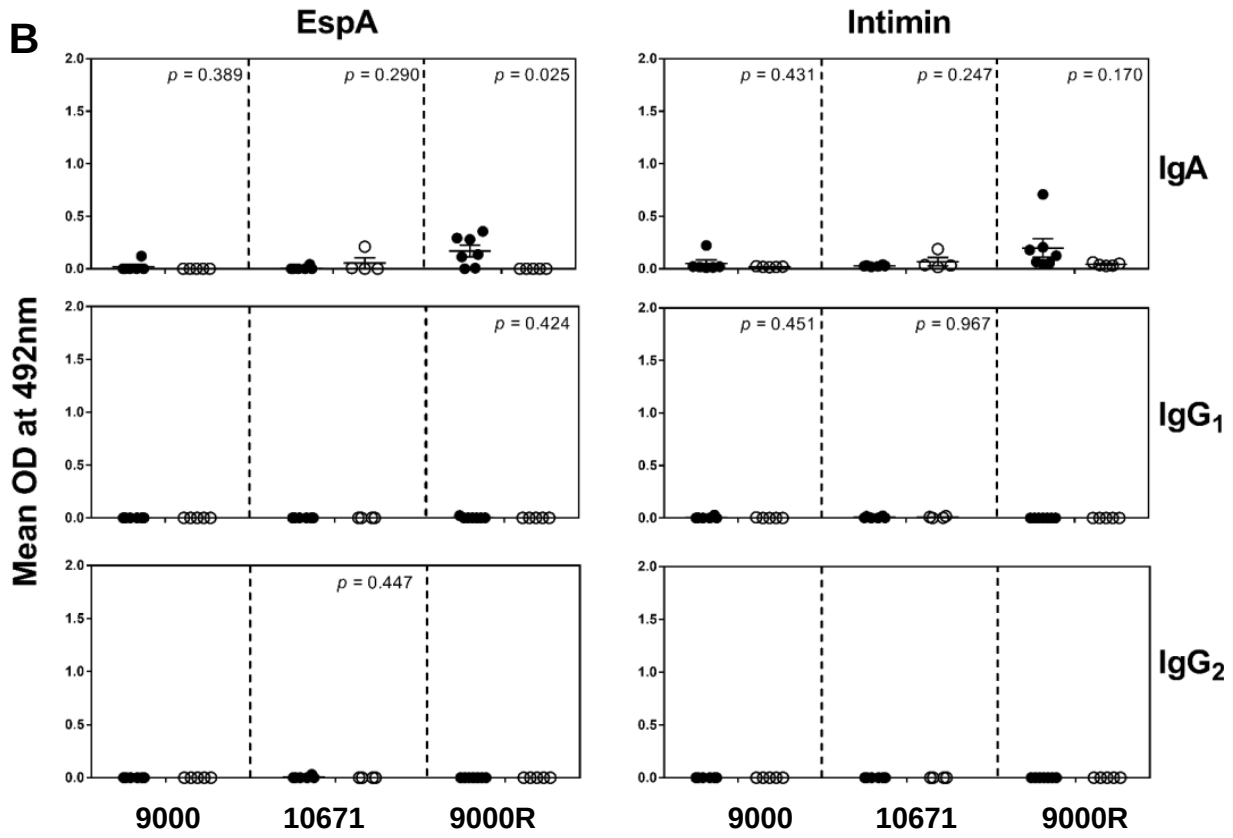


Figure 3.11. Levels of *E. coli* O157 specific antibodies within rectal lymph node antibody secreting cell probes (ASC) generated from *E. coli* O157 challenged and unchallenged calves. ASC probes were generated from rectal lymph nodes of calves challenged with strain 9000 (PT21/28 *stx*2c+), strain 10671 (PT32 *stx*2c+) or strain 9000R (PT21/28 *stx*2c+ *stx*2a+) (closed circles) and their relevant unchallenged controls (open circles). (A) Levels of H7 and Tir specific antibodies. (B) Levels of EspA and Intimin-specific antibodies. The error bars represent mean and standard error of the mean. Each symbol represents an individual animal. Unpaired t-tests were used to compare antibody levels between challenged and control calves for each strain and associated p values are indicated.

3.3.4. Contribution of Stx2a to survival within the ruminal gastro-intestinal tract (DO 1.3.3)

From our transmission studies, it was clear that the ability to produce a functional Stx2a was associated with higher levels of calf-to-calf transmission but this was not associated with suppression of immune responses to *E. coli* O157. One alternative explanation is that Stx2a may allow more efficient transit of bacteria through the rumen to the site of colonisation in the lower intestinal tract by killing rumen protozoa, which have been shown by other groups to predate on *E. coli* O157. This has been proposed to be a key reason why *E. coli* O157 has retained the capability of expressing Stx[25], although this theory has recently been called into question[75]. This would then enable strains capable of expression of functional Stx2a to colonise

the lower intestinal tract at lower doses, which are more likely to reflect those ingested by calves from their environment, which would in turn result in enhanced transmission rates.

To determine whether infectious dose and/or the ability to survive transit through the rumen was different between Stx2a+ and Stx2a- strains, we conducted a study in which calves were challenged by orogastric intubation with a low dose (10^3 cfu) of either strain 9000 (*stx*2c+) or the isogenic strain 9000R (*stx*2a+, *stx*2c+). This dose was selected as it reflected a potential dose that sentinel calves could have ingested from their environment based on environmental levels detected in pens used for the transmission studies. Challenged male Holstein-Friesian calves (n=3 per strain) were fitted with abomasal cannulae to allow sampling of abomasal fluid (representing ruminal outflow) in addition to faecal samples. Calves were conventionally reared between 9 and 10 weeks of age and fully weaned at the time of challenge, being fed on hay and calf concentrate for 3 weeks prior to challenge, and therefore had the opportunity to establish their ruminal flora. A different pen was used for each strain to avoid cross-contamination of samples with the two different strains. Samples were collected at 0hr, 6hr, 24hr and daily thereafter for three weeks and bacterial numbers quantified. Results from analysis of the abomasal fluid are shown in Table 3.3. Strain 9000 was detected following broth enrichment of abomasal fluid samples in 2/3 calves on days 3, 7 and 9 post-challenge, most likely reflecting oral re-sampling of the bacteria by the calves from the environment. Strain 9000R was not detected in the abomasal fluid at any time-point.

Table 3.3. Presence of *E. coli* O157 in abomasal fluid following low dose oral challenge

Calf	Strain	Days post-challenge																
		0	1	2	3	4	7	8	9	10	11	13	14	15	16	17	18	21	23
505047	9000	-	-	-	**+**		-	-	-	-	-	-	-	-	-	-	-	-	-
103654	9000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
105043	9000	-	-	-	-	-	**+**	-	**+**	-	-	-	-	-	-	-	-	-	-
505047	9000R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
103654	9000R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
105043	9000R	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

+ = detection via broth enrichment; - = no bacteria detected

Faecal bacterial excretion from the challenged calves is shown in Figure 3.12. For strain 9000, 2/3 calves became colonised. Both of these calves shed bacteria at > 10^4 cfu/g faeces. The challenge strain was detected sporadically following broth enrichment throughout the study period but were <33 cfu/g faeces at all time-points (33 cfu/g is the limit of detection for the direct plating assay). For strain 9000R, again 2/3 calves became colonised following challenge, with both shedding ≥ 10^4 cfu/g faeces. The third calf was positive by direct plating on only one occasion (day 21 post-challenge).

These results do not support the hypothesis that strain 9000R is better able to survive transit through the rumen than strain 9000, indicating that the higher transmission rates associated with Stx2a are unlikely to be due to reduced predation of *E. coli* O157 by ruminal protozoa.

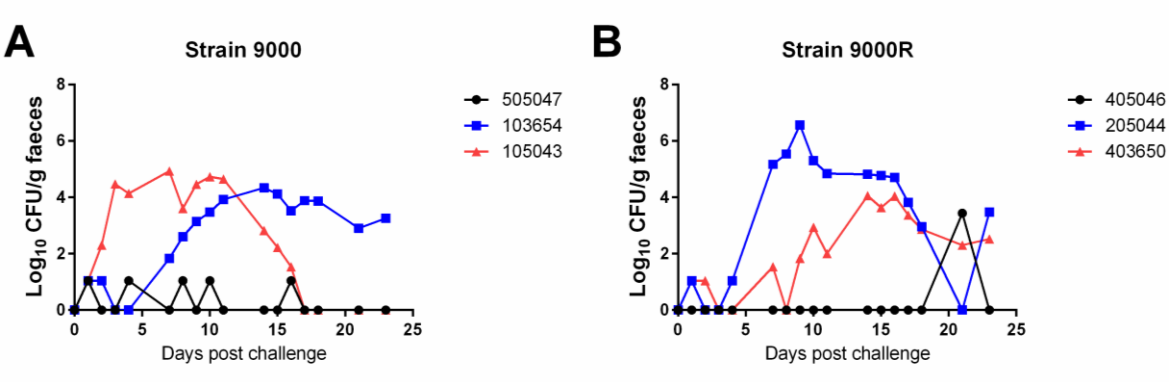


Figure 3.12. Excretion of *E. coli* O157 strains in calves following low dose oral challenge. Calves (n=3) were challenged with 10^3 cfu of either (A) strain 9000 (PT21/28 *stx*2c+) or (B) isogenic strain 9000R (PT21/28 *stx*2a+ *stx*2c+). Levels of bacterial in faeces were monitored for 23 days post-challenge.

3.4. Conclusions from Objective 2

Through controlled challenge studies we demonstrate that Stx2a is associated with enhanced cattle to cattle transmission of *E. coli* O157, as defined by the ability of the bacteria to transmit between cattle and efficiently colonise in-contact calves. Furthermore, as both PT21/28 isolates were excreted and transmitted at significantly higher levels than the PT32 isolate irrespective of their capacity to produce Stx2a, other characteristics of PT21/28 must contribute to higher excretion and transmission. While these characteristics are currently unclear, we have detected a number of genetic differences between the PT21/28 and PT32 strains, which could contribute to the enhanced excretion and transmission of the PT21/28 strains. *E. coli* O157-specific antibody responses were greatest for the PT21/28 strain capable of producing functional Stx2a, and possession of Stx2a did not appear to enhance bacterial survival through the rumen, although this is based on only one study. This suggests that the enhanced cattle-to-cattle transmission rates associated with Stx2a are not due to either down-regulation of local (mucosal) adaptive immune responses to the bacteria or enhanced killing of predatory protozoa within the rumen.

4. Objective 3: To test a vaccine using a super-shedding strain and model the impact of on-farm interventions based on data generated within the programme

4.1 SUMMARY

Three representative *E. coli* O157 strains (a wild type PT21/28 strain expressing Stx2c only, a wild type PT32 strain expressing Stx2c only, and an isogenic strain of PT21/28 capable of expressing both Stx2a and Stx2c) were used in challenge experiments to quantify their transmissibility and potential for super-shedding, which have previously been described in Objective 2. The strain with the highest transmission potential was then used to test the capacity of a vaccine to limit transmission. The data arising from these studies were used to parameterise an epidemiological model of *E. coli* O157 shedding and transmission. These models were then used to estimate the impact of a number of vaccine strategies on *E. coli* O157 prevalence in Scotland using models that capture spread within and between herds.

The main findings of this objective were:

R_0, the average number of secondary cases arising from a single infected case, was estimated to be around 4.5 for the isogenic PT21/28 strain, which is similar to the estimates of R_0 on field-derived faecal *E. coli* O157, indictating the relevance of using this strain for the vaccine efficacy study.

An experimental *E. coli* O157 subunit vaccine was able to limit transmission of the isogenic PT21/28 strain to naive in-contact calves under controlled pen trial conditions.

Transmission models fitted to the vaccine pen trial data suggested a median vaccine efficacy of 74%.

Models of transmission in Scottish cattle herds predicted that vaccination administered either annually or at the point of movement would be sufficient to reduce the between herd prevalence of *E. coli* O157 from ~25% to 0%, implying a significant impact on *E. coli* O157 prevalence on Scottish herds within 3-4 years of treatment.

4.2. Background to vaccine and modelling research

Super-shedders and disease control in Scottish herds: Previous modelling work using data from SEERAD and IPRAVE *E. coli* O157 prevalence studies indicates that most new infections of *E. coli* O157 in cattle are driven by "super-shedders", individuals that shed bacteria at much higher rates than regular individuals[38]. These models also predict that limiting shedding of *E. coli* O157 in cattle to below 10^5 or 10^4 cfu/g faeces can reduce cattle-to-cattle transmissions by 52% and 88% respectively[38]. Furthermore, modelling also predict that reductions in *E. coli* O157 shedding in cattle will result in substantially greater reductions in human cases of than would be predicted from reductions in frequency of shedding in cattle alone: for example, reducing the numbers of cattle shedding *E. coli* O157 by 50% is predicted to reduce the number of human cases by around 80%[41]. Together, these data suggest that control of *E. coli* O157 in cattle populations is achievable with relatively modest levels of reductions in bacterial shedding, and that control in cattle may translate to significantly lower number of human cases.

However, it is important to note that these model outputs were not validated experimentally, and were not parameterised with data from actual intervention trials aimed at reducing *E. coli* O157 shedding in cattle, which would provide a more realistic assessment of the effectiveness of currently available interventions. Furthermore, it did not take into account variation in the transmissibility of different *E. coli* O157 strains between cattle which is evident from Objective 2.

The aim of this study was to use pen trial data to quantify the efficacy of a vaccine at preventing *E. coli* O157 spread in cattle. This vaccine efficacy would then be used within a standard modelling framework to quantify the expected impact at controlling *E. coli* O157 spread in Scottish herds.

Selected *E. coli* O157 strains: two different strains (PT21/28 strain 9000 and PT32 strain 107671) were chosen to be examined as they are commonly found in the Scottish cattle population. The PT21/28 strain we used, which is associated with super-shedding, was also modified by repairing the *stx*2a gene (strain 9000R), as this is thought to be important in transmission. The strain which shows the greatest potential to spread is also likely to be the most difficult to control. Therefore, two rounds of experiments were performed, the first to quantify transmission potential in the different strains (as described in Objective 1), and the second to quantify the efficacy of the vaccine in limiting calf-to-calf transmission with a highly transmissible strain (i.e. strain 9000R).

Vaccination as a feasible on-farm approach to control *E. coli* O157 in cattle: as detailed in Section 1.6, there are a number of on-farm interventions which have been explored to control *E. coli* O157 in cattle populations. Of the interventions tested, only probiotics and vaccines have been shown to be consistently effective at

reducing shedding of *E. coli* O157 in cattle[49]. However, while vaccines typically last several months or more, probiotics need to be administered on a daily basis[50], meaning they are less practical to deploy in Scottish cattle populations, particularly to grazing cattle. For these reasons, this objective focused on testing and modelling the impact of vaccine interventions to reduce *E. coli* O157 in cattle.

4.3 Materials & Methods

4.3.1. Data collection

The transmission experiments, performed as part of Objective 3 at Moredun Research Institute in Containment Level 3 pens, orally challenged six calves to a high dose (~10^9 cfu) of each *E. coli* O157 strain, then moved two of the infected calves into two other pens, each containing five susceptible calves (see Objective 3 and Fig. 3.5 for details). Both the number of bacteria shed by each calf (measured in cfu/g), and the level of contamination in the environment were recorded for the following 30 days.

The isogenic PT21/28 strain (9000R) showed the highest levels of super-shedding and the highest potential to transmit infection, and so it was chosen for the subsequent set of vaccine trials, which replicated the initial setup, but now included six calves, three of which were inoculated with an experimental subunit *E. coli* O157 vaccine.

4.3.2. Vaccination trial (DO 3.1.1)

The experimental design is illustrated in Fig. 4.1. Six conventionally reared male Holstein-Friesian calves negative for *E. coli* O157 by both IMS for *E. coli* O157 (using *E. coli* O157 Dynabeads; ThermoFisher) and quantitative PCR for *E. coli* O157:H7 *rfb* gene and *stx1* and *stx2* were immunised via the subcutaneous route with an experimental *E. coli* O157 subunit vaccine consisting of 60 µg recombinant EspA, 60 µg recombinant Intimin and 60 µg purified H7 flagellin plus 5 mg of the adjuvant Quil A (Brenntag-Biosector). The average age at vaccination was 55 days (range = 52-57 days). This vaccine was co-developed by Moredun Research Institute, the University of Edinburgh and SRUC and is highly effective at reducing shedding of *E. coli* O157 shedding in calves orally challenged with a high dose (10^9-10^{10} cfu) of bacteria[51,54] but is not currently licensed for commercial use. The three vaccine components are highly conserved, suggesting the vaccine will be effective against a wide range of *E. coli* O157 strains.

A further six calves were immunised with 5mg Quil A alone to act as adjuvant only controls. Immunisations were repeated on a further two occasions three weeks apart. One week after the final immunisation calves were moved into the MRI high security unit rooms C2 and C3, with each room containing three vaccinated and three adjuvant control calves. One calf shedding >10^4 cfu/g faeces of strain PT21/28 strain 9000 (generated by oral challenge with ~10^9 cfu of the strain in room C1 five days previously) was introduced into each room and transmission from the super-

shedding calf to in-contact vaccinated and control calves was monitored for a 24 day period.

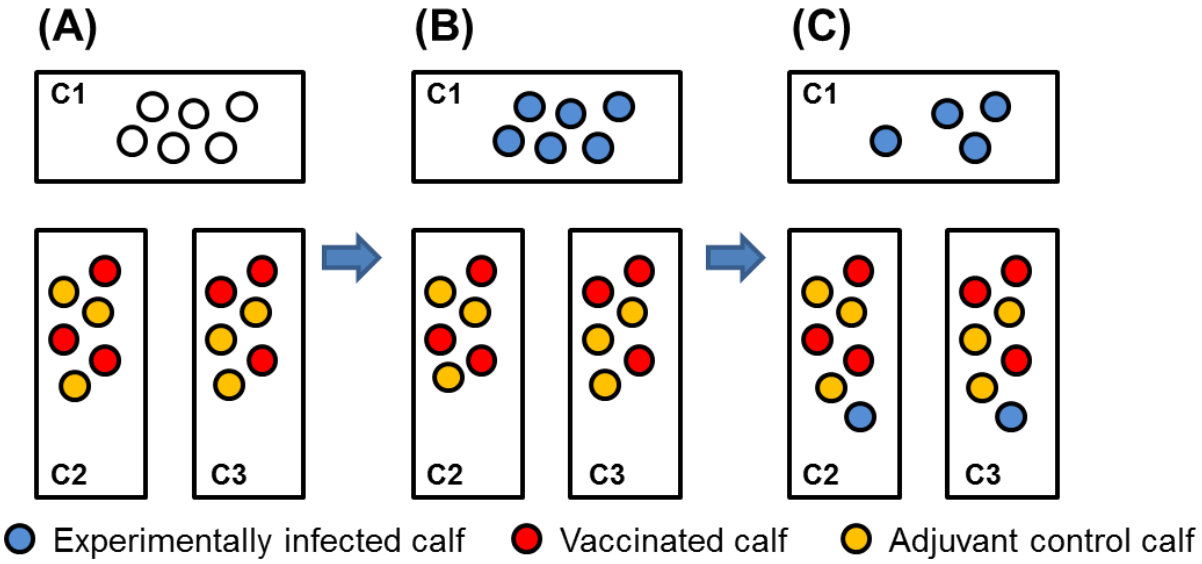


Figure 4.1. Experimental design of vaccination studies. *E. coli* O157 naïve calves were housed in three separate rooms (C1, C2 and C3) within the MRI High Security Unit as detailed in (A). Calves in C2 and C3 had either been previously vaccinated with an experimental *E. coli* O157 vaccine or adjuvant only control. All calves in C1 were orally challenged by orogastric intubation with ~10^9 *E. coli* O157 (B). Five days later, a calf shedding high levels of *E. coli* O157 (>10^4 cfu/g faeces) was moved into C2 and C3 (C). Faecal bacterial shedding from all calves, as well as levels of bacteria on the floors of each pen, were monitored for a further 19 day period.

4.3.3. Modelling approach (DO 3.2.1)

Background details of the modelling used in this objective are detailed in APPENDIX F. Specific information on the models used in this objective is shown below.

4.3.3.1 SLHv-R model

Calves are categorised as either susceptible calves (shedding 0-29 cfu/g faeces), low shedders (shedding 30-9,999 cfu/g faeces), or high shedders (shedding >10,000 cfu/g faeces), represented by the variables S, L, and H respectively. In addition, vaccinated calves for each category are represented by the variables S_v, L_v, and H_v. The level of environmental contamination, the reservoir, is represented by the variable R.

Susceptible calves become low shedders via exposure to the reservoir at rate $\beta_{SL}RS/N$ (where N is the number of calves in the model), and low shedders progress

to high shedders at rate $\beta_{LH}L$. High shedders recover to low shedding at rate $\alpha_{HL}H$, and low shedders recover to susceptible at rate $\alpha_{LS}L$. Low and high shedders shed into the reservoir at rate $\gamma_L L$ and $\gamma_H H$ respectively, and the reservoir decays at rate λR.

Vaccinated calves are subject to the same processes as unvaccinated calves, however we assume that the colonisation rate is reduced by the vaccine efficacy p_v (i.e. $p_v = 1$ means all colonisation is blocked, and $p_v = 0$ means no colonisation is blocked). The main effect of the vaccine is expected to be a reduction in the number of individuals which shed high levels of *E. coli* O157 (>10^4 cfu) rather than preventing individuals shedding low levels of bacteria. This is because the vaccine targets proteins associated with adherence to the intestinal epithelium[51,54] and high shedding is associated with epithelial adherence. In contrast, low levels of shedding are more likely to be reflect bacteria transiting through the gastrointestinal tract without adhering to the gut epithelium, which would be minimally affected by the vaccine. See Fig. 4.2 for a model schematic and Table 4.1 for a description of the model terminology.

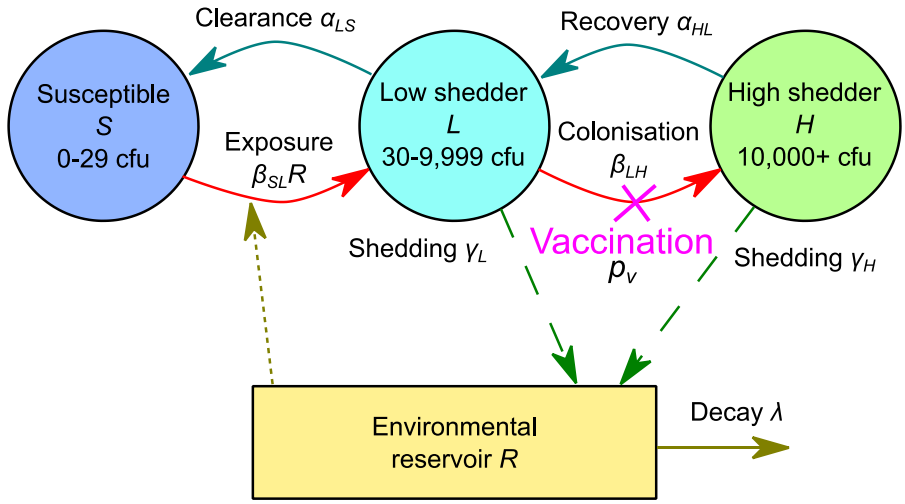


Figure 4.2. Schematic of the SLHv-R model. See Table 4.1 for parameter estimates.

Table 4.1. Terminology of the SLHv-R model

Parameter	Symbol
Infection rate from *S* to *L*	β_{SL}
Promotion rate from *L* to *H*	β_{LH}
Recovery rate from *H* to *L*	α_{HL}
Clearance rate from *L* to *S*	α_{LS}
Shedding rate for *L* into *R*	γ_L
Shedding rate for *H* into *R*	γ_H
Shedding variance	γ_V
Decay rate for *R*	λ
Vaccination efficacy	p_v

Given the SLHv-R model, R_0 may be calculated using the method described by Diekmann & Heesterbeek[38] using only the parameters of the model:

$$R_0 = \frac{\beta_{SL}\,[\,(\mu + \alpha_{HL})\gamma_L + \beta_{LH}\,\gamma_H]}{\lambda\,[\mu\,\beta_{LH} + (\mu + \alpha_{LH})(\mu + \alpha_{HL})]}$$

This allows for R_0 to be calculated as part of the data fitting process.

4.3.3.2 Fitting the model to the data using Markov chain Monte Carlo (MCMC)

To fit the model to the data, each animal was categorised as either “Susceptible” if it shed between 0-29 cfu/g, “Low shedder” if it shed 30-9,999 cfu/g, and “High shedder” if it shed >10,000 cfu/g. The 30 cfu/g threshold was chosen as the direct plating test was only sensitive to levels of about 100 cfu/g (see Table 3.2 in Objective 2), but was repeated 3 times, and so if 1 in 3 tests was positive, this would give an average of 33 cfu/g. Choosing the threshold to be 30 ensured that this was counted as a low shedder. Missing data were treated as latent variables, and MCMC was used to obtain posterior distributions for each model parameter.

The data is in the form of daily measurements, to which a continuous time model was fitted. Given a parameterised continuous time model, instantaneous rates of $X \to Y$ transitions can be specified by a rate matrix Q, and then the probability of a transition sometime between the interval $[t, t+1)$ may be calculated using the matrix exponential $P = \exp(-Qt)$, where $t = 1$. The matrix P accounts for all possible transitions that may occur in the time interval, which also has the side effect of allowing for $S \leftrightarrow H$ transitions in the model, allowing any variable to be used as a latent variable when data is missing. The matrix exponential is not a suitable method for calculating the transition from $R(t) \to R(t+1)$, since there are too many possible states. Instead, we calculate the probability of the transition by considering the new contribution from low and high shedders, and the decay of the existing

contamination. The reservoir $R(t+1)$ is therefore assumed to be normally distributed with

- mean: $\gamma_L L(t) + \gamma_H H(t) + \exp(-\lambda) R(t)$, and
- variance: γ_v

allowing the probability to be calculated.

Removing the reservoir

We have previously demonstrated that an SLHv model with no reservoir could behave similarly to the SLHv-R model for parameters similar to the ones obtained from the transmission experiments, provided the transmission rate was adjusted from:

$$\beta_{SL} RS/N$$

to

$$\frac{\beta_{SL}}{\lambda} (\gamma_L L + \gamma_H H) S/N$$

This allows for removal of the reservoir variable *R*, simplifying modelling in the metapopulation model.

4.3.3.3 Metapopulation model

Matthews *et al*.[38] calculated R_0 estimates for *E. coli* O157 on Scottish cattle herds. The values obtained from the transmission experiments are in considerably different conditions (6-7 calves in a containment level 3 pen versus a larger herd outside in a field), and so estimates for R_0 may be different. Obtaining estimates for the super-shedding dynamics and vaccine potential is still an important result, and these can be used in a metapopulation model but adjusting the transmission coefficient to obtain an R_0 that matches the Matthews *et al*. results. A schematic of the cattle movements between herds is shown in Figure 4.3.

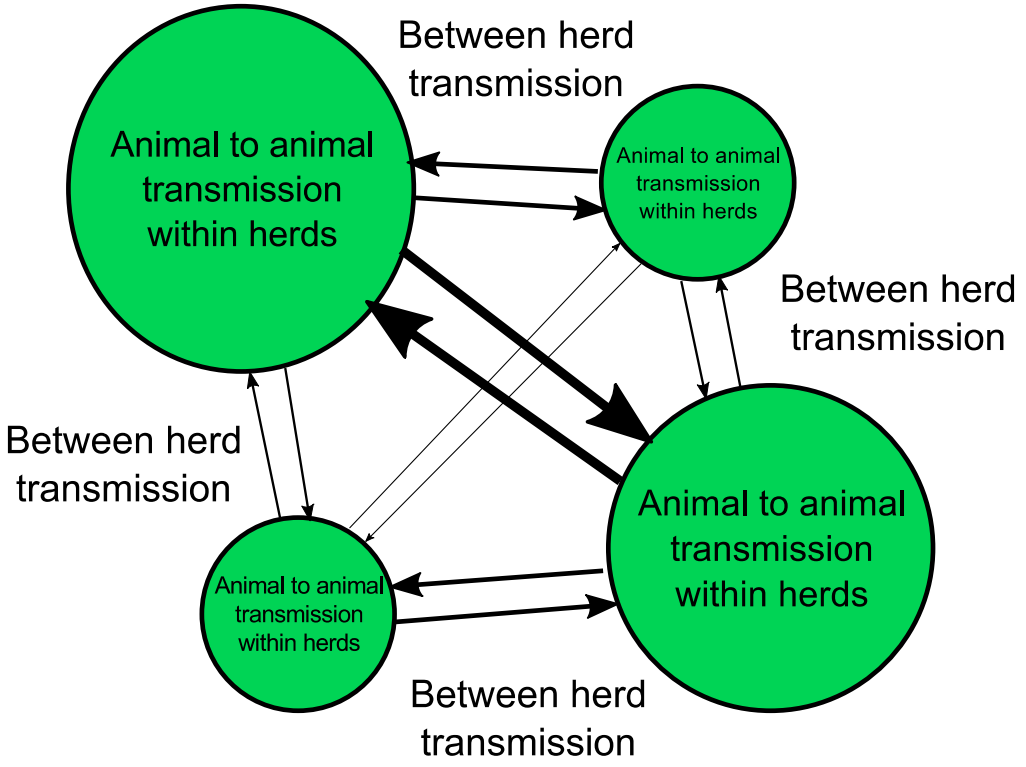


Figure 4.3: Schematic of cattle movements between groups. Movement in and out is proportional to herd size, and consequently larger herds form stronger connections with the rest of the population. The thickness of the arrows represent the level of transfer between groups.

Using data on Scottish cattle herd sizes (mean 18.9, standard dev 15.1, range 1-140), a set of herds were randomly generated with a similar distribution of herd sizes, and the population was seeded with infectives.

Dynamics within each herd used the SLHv model, but individuals were allowed to move from one herd to another (randomly chosen, but weighted by herd size, so movement in = movement out) at *per capita* rate κ. Mortality was introduced at *per capita* rate μ, removing infected and vaccinated animals. New animals were introduced into the susceptible class at constant rate μN_i, so herd sizes fluctuate around N_i.

In addition, waning immunity to vaccination was introduced. A similar vaccine based on T3S system-associated proteins from *E. coli* O157 is estimated to last for 2-3 months[38], although this period was too long to be confirmed in the transmission experiments conducted. To be conservative, we selected the lower bound of 2 months. Therefore, vaccine waning at *per capita* daily rate of $\rho = 0.0164$ was included in the model.

The model, which captures the epidemiology as well as cattle population demography, was allowed sufficient time to reach a quasi-equilibrium (where the proportion of infected herds was not rapidly changing), which took about 4 years,

and then disease intervention was introduced, with a variety of intervention scenarios being explored.

4.4. Results

4.4.1. Data collection

Individual shedding data from naïve sentinels exposed to calves super-shedding either PT21/28 strain 9000, PT21/28 strain 9000R or PT32 strain 10671 originating from transmission studies in Objective 2 are shown graphically in Fig. 4.4.

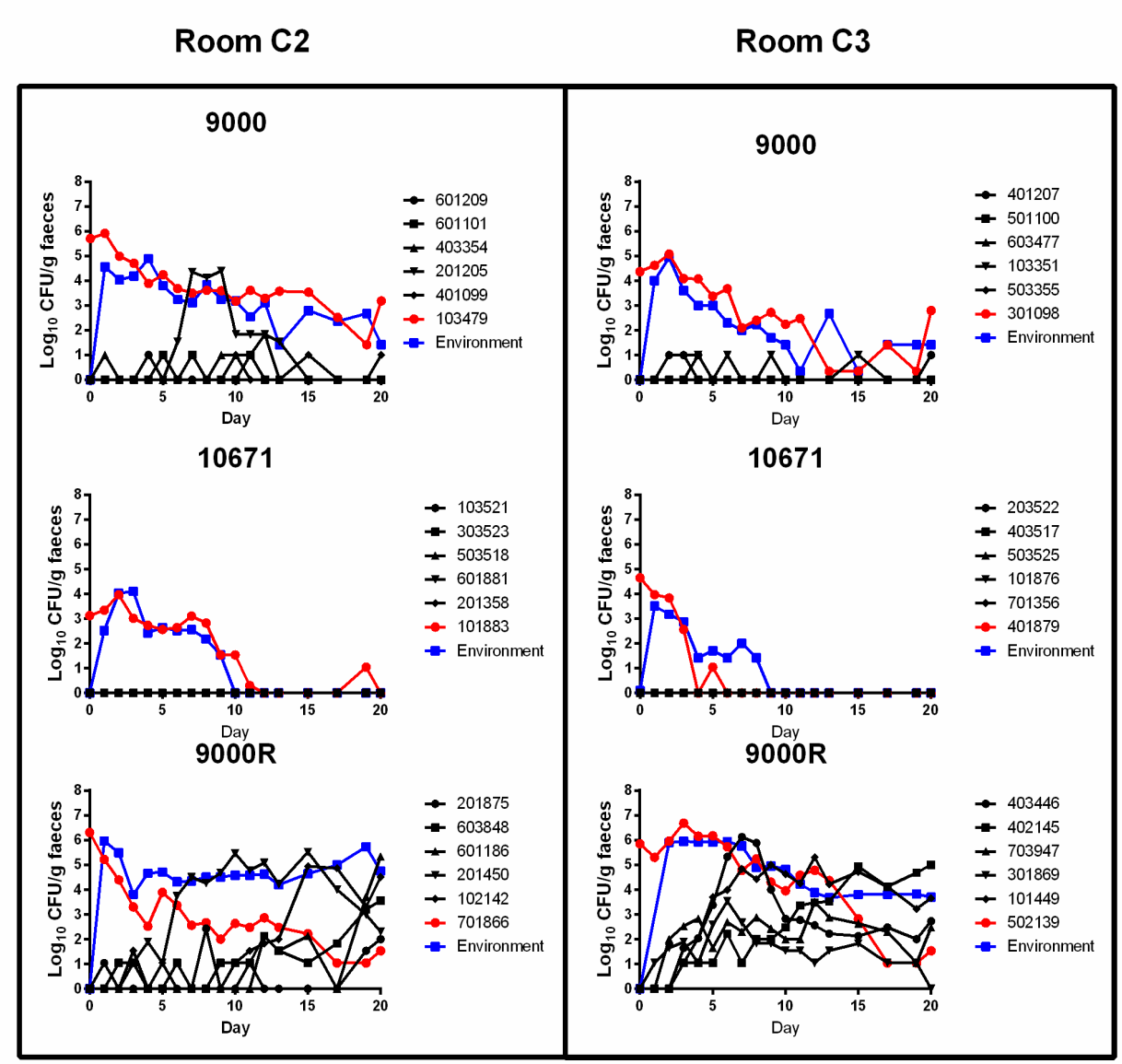


Figure 4.4. Individual shedding curves for calves colonised with *E. coli* O157 strains 9000, 10671 and 9000R. Bacterial shedding was monitored from experimentally challenged animals (red) and naïve sentinel calves (black) in rooms C2 and C3 for three strains of *E. coli* O157 (9000, 9000R and 10671). Environmental levels within each room were also recorded and are indicated in blue. Data points represent the cfu/g faeces (for calves) or per gram of environmental material (for environmental levels) from three replicate plate counts.

4.4.2. Vaccine trial data (DO 3.1.1-3)

The results of mean bacterial shedding of vaccinated and control calves in rooms C2 and C3 are shown in Fig. 4.5. All calves in contact with the introduced super-shedding calves became colonised; however, peak shedding was lower in the vaccinated calves, with no calves (6/6) shedding >10^3 cfu/g faeces. In contrast, in the adjuvant only control group 3/6 calves shed >10^3 cfu/g, and 2/6 calves shed >10^4 cfu/g faeces at least one time-point. Total bacterial shedding, as determined by AUC analysis, was significantly lower in vaccinated calves compared to adjuvant only controls ($p = 0.031$); however mean shedding levels over time were not significantly different between vaccinated and control calves (repeated measures analysis $p = 0.078$). Antibody responses to the vaccine antigens confirmed that all calves seroconverted to the vaccine (Fig. 4.6).

This data provides evidence that the vaccine had a protective effect in reducing transmission of *E. coli* O157 from super-shedding to naïve calves, and prevented in-contact calves becoming 'super-shedders' (i.e. shedding >10^4 cfu/g faeces). It should also be noted that by mixing vaccinated and not-vaccinated calves rather than vaccinating all calves in each pen (which would be optimal to achieve 'herd immunity'), and by using a highly transmissible strain of *E. coli* O157, this study may underestimate the effectiveness of the vaccine in the field.

The data is not comparable to previous vaccination studies[52,54] as these involved oral bacterial challenge of vaccinated and control calves rather than assessing the potential of the vaccine to block transmission.

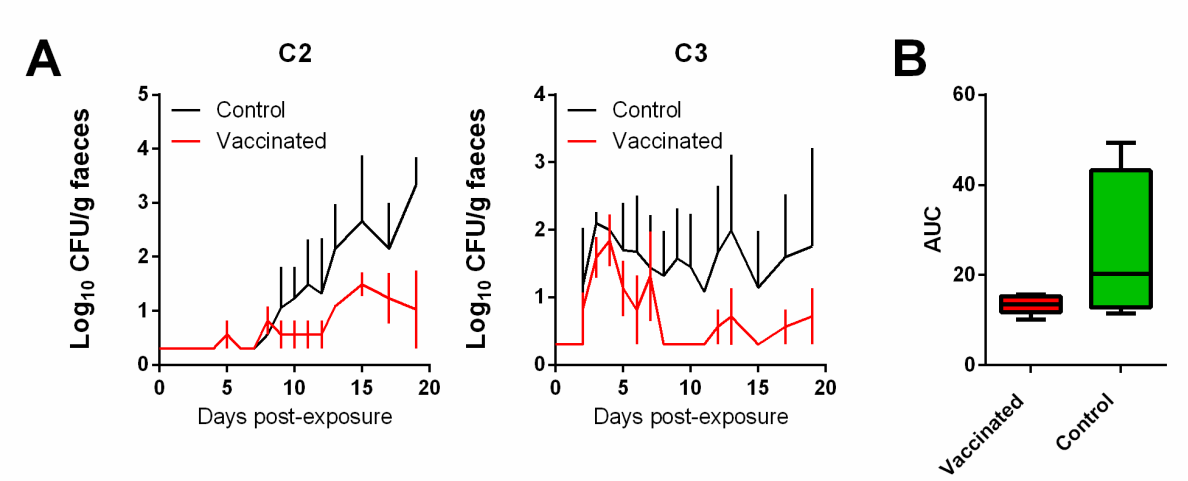


Figure 4.5. Mean and total shedding of *E. coli* O157 from calves vaccinated with the experimental *E. coli* O157 vaccine and exposed to *E. coli* O157 super-shedding calves. Calves either vaccinated with an experimental *E. coli* O157 vaccine (vaccinated) or adjuvant only (control) were exposed to calves shedding >10^4 cfu of *E. coli* O157 strain 9000R. The study was performed in two pens, C2 and C3, each of which contained three vaccinated and three control calves and one calf shedding *E. coli* O157. (A) Shedding levels of strain 9000R from vaccinated and control calves were monitored over a 19-day period in pen C2 or C3. Data represents the mean ± SE of the mean. (B) Area under the curve data for vaccinated and control calves over the duration of the study.

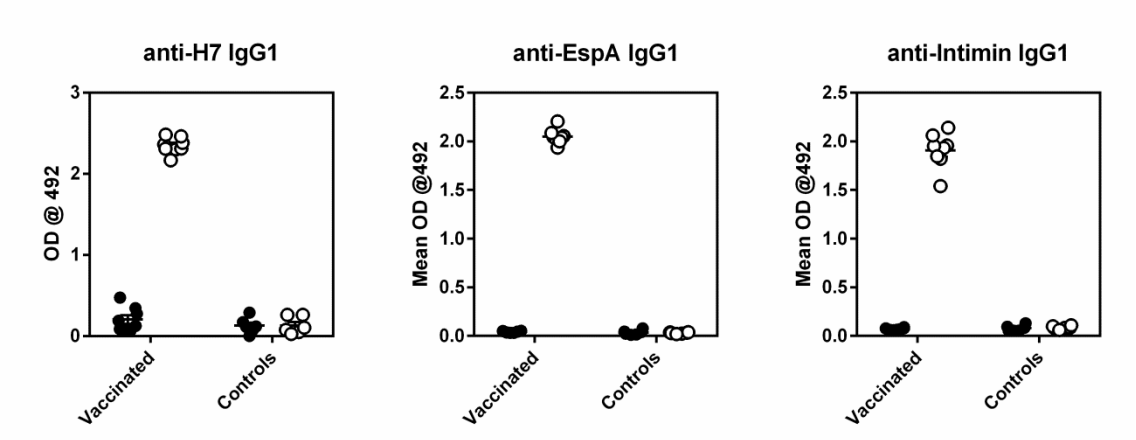


Figure 4.6. Serum antibody responses in calves vaccinated with the experimental *E. coli* O157 vaccine. Levels of IgG_1 specific to the vaccine antigens H7, EspA and Intimin were determined by ELISA prior to the first vaccination and 1 week after then final vaccination in calves immunized with the *E. coli* O157 vaccine or adjuvant only controls. Antibody levels are indicated by Optical Density readings at 492 nm (OD @492) which is directly proportional to the amount of antibody within the sample.

4.4.3. Model parameters (DO 3.2.1)

Models were used to estimate parameters for sentinel calves exposed to calves infected with either wild-type PT21/28 strain 9000, isogenic PT21/28 strain 9000R or wild-type strain 10671, and to estimate vaccine efficacy for PT21/28 strain 9000R only. Given a cut-off point of 10^4 cfu/g between low shedders (L) and high shedders (H), median values and 95% credible intervals obtained via MCMC are given in Table 4.2.

Table 4.2. Median and 95% credible intervals of parameter estimates obtained via MCMC. A single estimate is provided for the shedding variance and the decay rate of the reservoir as these were assumed constant across strains. Vaccine trials were not conducted for the Wild type PT21/28 strain 9000 and Wild type PT32 strain (indicated by N/A for Vaccine efficacy).

Parameter	**Symbol**	**Wild type PT21/28 strain 9000**	**Isogenic PT21/28 strain 9000R**	**Wild type PT32 strain 10671**
Infection rate from *S* to *L*	β_{SL}	3.2×10^{-5} (1×10^{-5} - 5×10^{-5})	1.9×10^{-5} (1×10^{-5} - 2.7×10^{-5})	2.9×10^{-5}
Promotion rate from *L* to *H*	β_{LH}	0.068 (2×10^{-3} - 0.20)	0.12 (0.068-0.19)	0.39 (3×10^{-4} - 2.1)
Recovery rate from *H* to *L*	α_{HL}	0.297 (5.9×10^{-2} - 0.67)	0.248 (0.13 - 0.4)	7.11 (2.4 - 10.0)
Clearance rate from *L* to *S*	α_{LS}	0.283 (0.1 - 0.5)	0.258 (0.17 - 0.36)	0.219 (0.02 - 0.5)
Shedding rate for *L* into *R*	γ_L	14.7 (10 - 36)	185 (16 - 780)	25 (10 – 120)
Shedding rate for *H* into *R*	γ_H	1.45×10^{4} (1×10^{4} - 2.7×10^{5})	1.20×10^{5} (8.4×10^{4} - 1.6×10^{5})	1.15×10^{5} (1×10^{4} - 6.9×10^{5})
Shedding variance	γ_v	1.63×10^{5} (1.1×10^{5} - 2.3×10^{5})		
Decay rate for *R*	λ	1.38 (0.93 – 2.2)		
Vaccination efficacy	p_v	N/A	0.744 (0.14 - 1.0)	N/A

Three parameters of interest when determining the risk of a calf becoming a super-shedder (and thus contribute significantly to calf-to-calf transmission) are the rate of promotion to (β_{LH}), and recovery from (α_{HL}), super-shedder status (since these are the parameters that characterise a strain's propensity to super-shedding), and the vaccine efficacy (p_v). The posterior distributions for these parameters, obtained from the MCMC are shown in Figs 4.7-4.9.

Figure 4.7 shows the posterior distribution of the rate of promotion to super-shedder status and indicates that calves infected with the isogenic PT21/28 strain 9000R are more likely to be promoted to super-shedder status than those infected with the wild type PT21/28 strain 9000R.

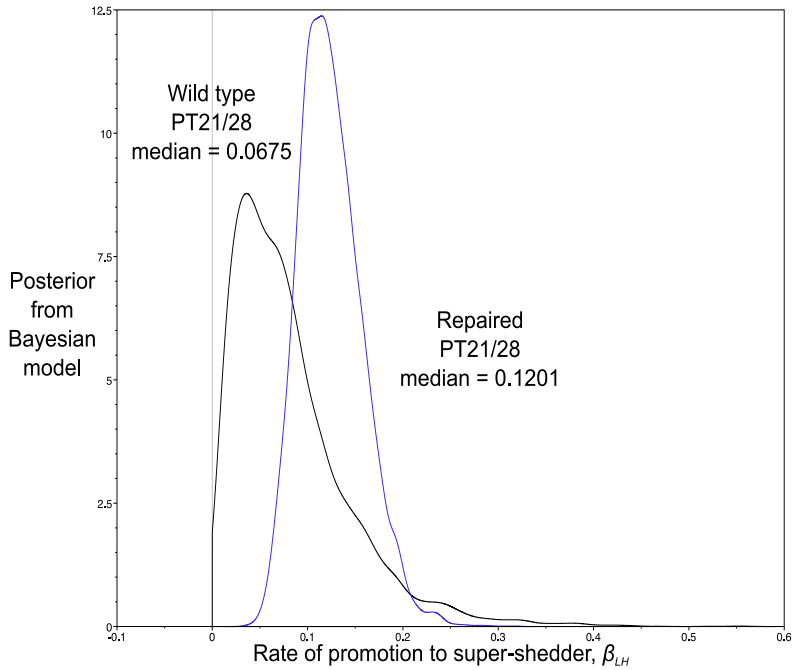


Figure 4.7. Posterior distribution of the rate of promotion to super-shedder, β_{LH}. Wild type PT21/28 = strain 9000; Repaired PT21/28 = strain 9000R.

Figure 4.8 shows the posterior distribution of the rate of recovery from super-shedder status. The PT21/28 strain 9000R has a lower mean value compared to the wild-type PT21/28, suggesting a longer period as a super-shedder before recovery. This, combined with the higher rate of promotion to super-shedder, leads to increased prevalence of super-shedders with strain 9000R.

Posterior distribution of the vaccine efficacy is shown in Figure 4.9. This provides evidence that the vaccine is effective, with a median probability of preventing infection being 0.74, or 74%, with a 95% credible interval of (0.08-0.99) indicated by the region in blue in Figure 4.9.

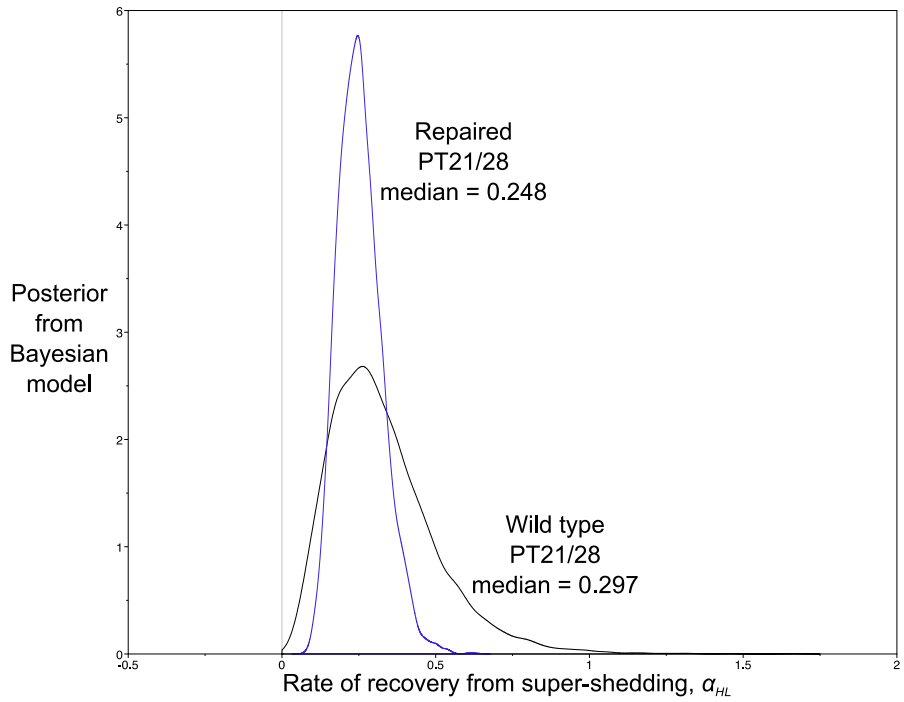


Figure 4.8. Posterior distribution of the rate of recovery from super-shedder status, α_{LH}. Wild type PT21/28 = strain 9000; Repaired PT21/28 = strain 9000R

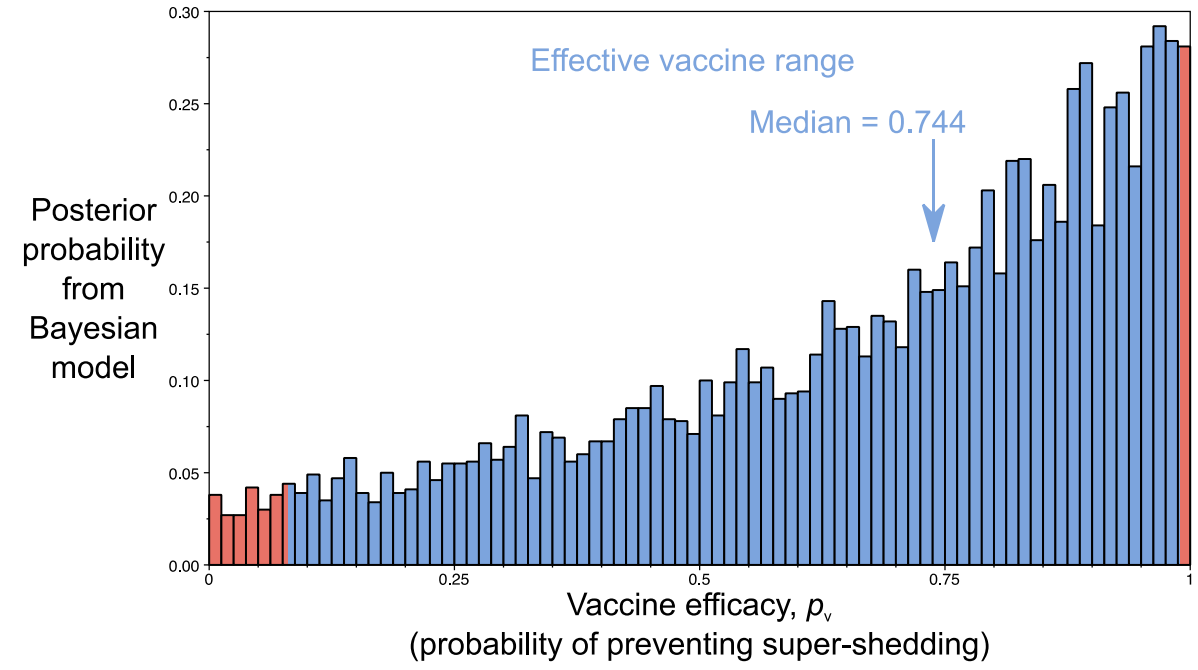


Figure 4.9. Posterior distribution of the vaccine efficacy, p_v. The 95% credible interval (the Bayesian counterpart of a confidence interval from frequentist statistics) excludes $p_v = 0$, providing evidence against the null hypothesis of an ineffective vaccine and strong evidence for an effective vaccine.

To explore the effect of the cut-off points between low and high shedding on the predicted vaccine efficacy, we compared predicted vaccine efficacy across a number of different cut-off points from 2 x 10^2 to 1 x 10^6 cfu/g (Figure 4.10). This showed that predicted vaccine efficacy was similar across a range of cut-offs.

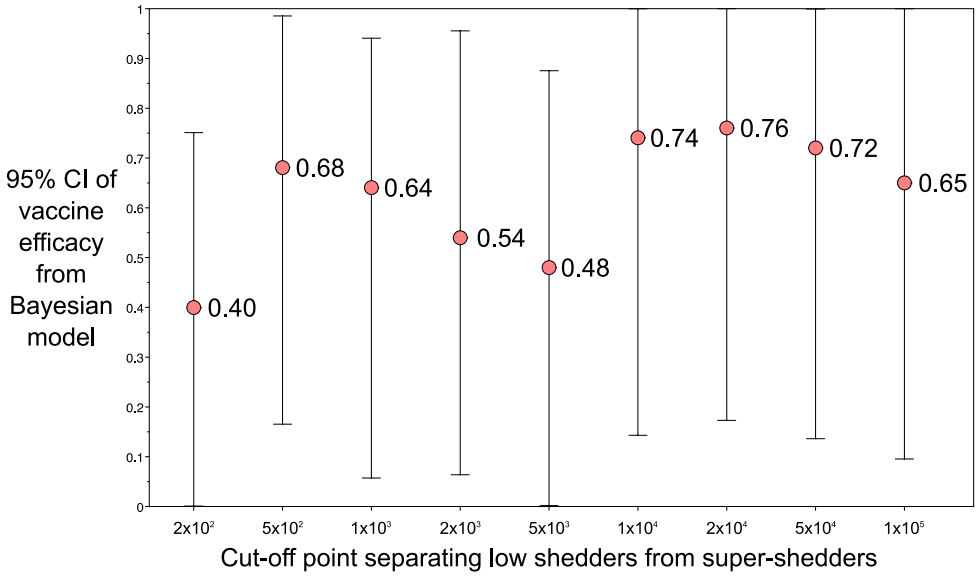


Figure 4.10. Median values (red dots) and 95% credible intervals of the vaccine efficacy, $\boldsymbol{p}_v$, for different values of the cut-off point separating low shedders and super-shedders.

Finally we determined the effect of using different cut-off points between low and high shedders on predictions of R_0 for each of the three *E. coli* O157 strains. For the isogenic PT21/28 strain, R_0 was fairly consistent, ranging from $R_0 = 4.2$ for a cut-off of 200 cfu, to $R_0 = 2.3$ for a cut-off of 5x10^4 cfu. For the other two strains, R_0 varied considerably, suggesting insufficient data for those strains and a poor model fit. The lack of data for PT21/28 strain 9000 and the PT32 strain can be explained by the lower transmission rates for these strains (in the case of the PT32 strain no transmissions were observed), which resulted in less information for parameterising the corresponding models.

As the vaccine targets antigens which are highly conserved between different *E. coli* O157 strains, we predict that the vaccine would be effective against the majority of *E. coli* O157 strains, including those, such as strain 9000R, with high transmission potential.

4.4.4. Metapopulation simulations (DO 3.2.2-4)

Since the posterior distributions show that vaccine efficacies p_v can lie within a wide range of values, we must consider how the different possible outcomes depend on that distribution. Therefore we ran simulations for a range of p_v corresponding to the [2.5, 25, 50, 75 and 97.5] percentiles i.e. p_v = [0.08, 0.5, 0.74, 0.89, 0.99] and vaccination at the point of movement (i.e. individuals that move between groups are

simultaneously given the vaccine), and annual vaccination are shown in Fig. 4.11 and Fig. 4.12 respectively. Since transmission rates in a pen may be higher than those in the field, we rescaled the transmission parameter β_{SL} so that R_0 reflects field conditions[38], while still accounting for the super-shedding dynamics observed in the pen trial. In all cases, vaccination reduced the overall prevalence and the proportion of infected herds, and for p_v = 0.74 (the median value from the posterior distribution, Figure 4.9), substantial reductions were found.

Specifically, starting with an initial between-herd prevalence of approximately 28%, then after 3 years of intervention, based on the vaccine efficacy posterior the [2.5, 25, 50, 75, 97.5] percentile between-herd prevalences are [23.0, 4.0, 0.26, 0.022, 0.0] for annual vaccination, and [22.0, 4.8, 1.2, 0.76, 0.34] for vaccination on movement. This corresponds to a median reduction of 99.1% (95% CI: 17.9 – 100%) for annual vaccination and 95.7% (95% CI: 21.4 – 98.8%) for vaccination at the point of movement.

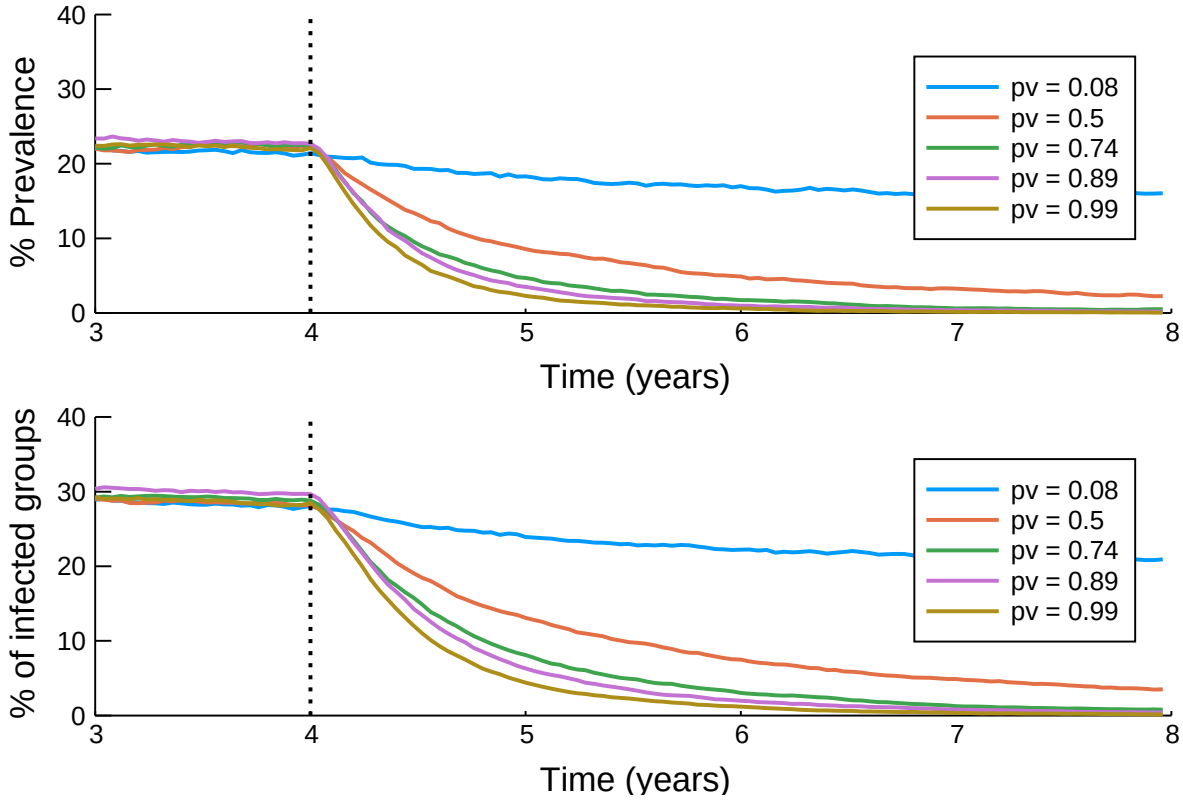


Figure 4.11. Disease control implemented by vaccinating individuals at the point of movement. Mean prevalence within infected groups (top) and proportion of infected groups (bottom) in the metapopulation model for p_v = 0.08 (blue), 0.5 (red), 0.74 (green), 0.89 (purple) and 0.99 (brown). The model is given four years to reach a quasi-equilibrium, then disease control (vaccination) is implemented. In all cases disease prevalence is decreased, and for $p_v \geq 0.74$ the infection is substantially reduced within three years.

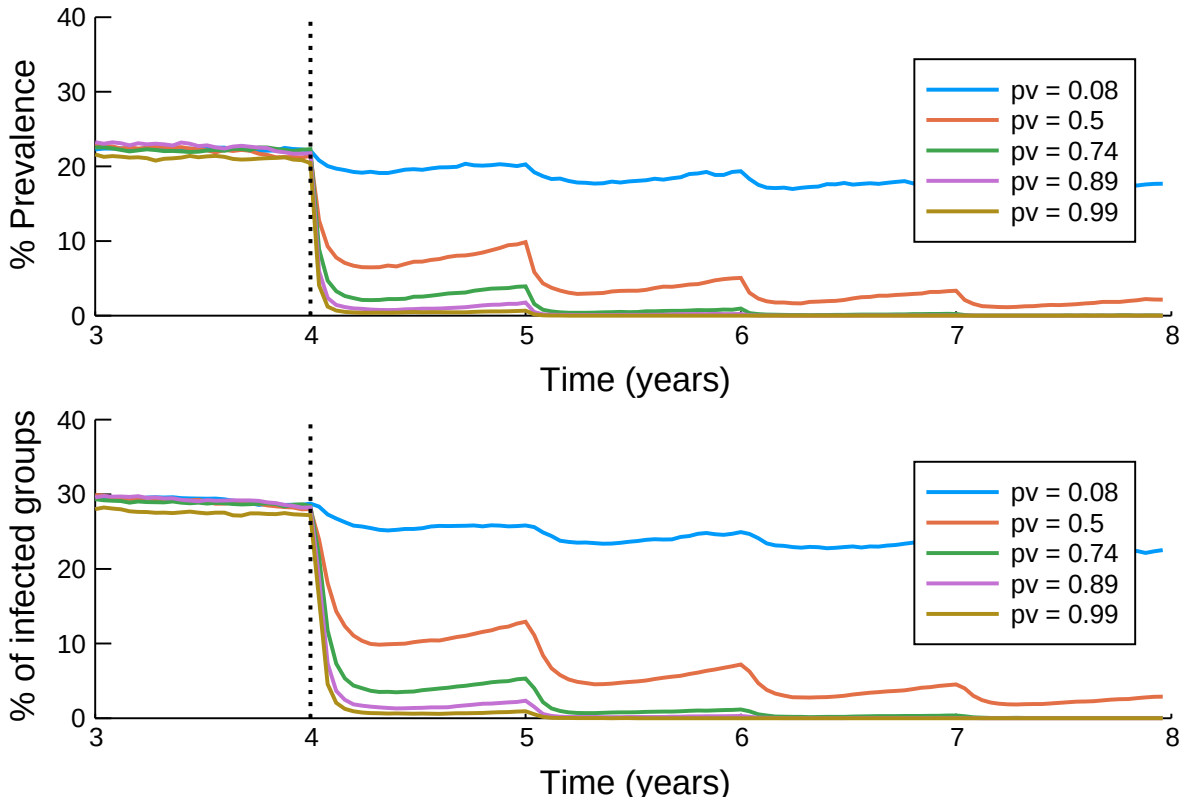


Figure 4.12. Disease control implemented by vaccinating individuals annually. Mean prevalence within infected groups (top) and proportion of infected groups (bottom) in the metapopulation model for p_v = 0.08 (blue), 0.5 (red), 0.74 (green), 0.89 (purple) and 0.99 (brown). The model is given 4 years to reach a quasi-equilibrium, then disease control (vaccination) is implemented. In all cases disease prevalence is decreased, and for $p_v \geq 0.74$ the infection is substantially reduced within three years.

4.5. Conclusions from Objective 3

The results of the challenge / vaccine trials strongly suggest that the vaccine is effective at reducing *E. coli* O157 transmission, and that by vaccinating individuals at the point of movement or via annual vaccination, the prevalence of even the most transmissible strains should be substantially reduced within three years.

However, the low numbers of individuals in the trials led to wide credible intervals on some of the parameter estimates, and there were insufficient super-shedding events to investigate more complex models of super-shedding dynamics. Also, the conditions in the Containment Level 3 pens differed significantly from the field conditions in which *E. coli* O157 normally persists. It is unknown whether this would affect the rate of super-shedding, but it is likely to affect the rate of transmission. It should also be noted that in the pen trials not all calves within each pen were vaccinated, and it is likely that the effect of the vaccine would be greater if this had been the case.

The next key step is to test this vaccine under field conditions. However, to do this will require the vaccine to obtain an appropriate license for use in animals destined for the human food chain, the requirements of which may differ depending on the country in which the vaccine will be tested. In the UK, this would require an Animal Test Certificate (ATC) from the Veterinary Medicines Directorate, which would require an appropriate safety testing of the vaccine in the appropriate age of cattle for which the vaccine was intended for use.

In summary, the results indicate that the tested vaccine could be highly effective in the field, but that a follow-up trial with larger numbers in field conditions could strengthen the confidence in the vaccine efficacy and improve knowledge of the super-shedding dynamics within hosts. Obtaining the appropriate license to test the vaccine under field conditions is now the critical next step.

5. Outputs

5.1. Recommendations

5.1.1 Knowledge exchange to reduce the risk to humans

A principal message from the results of the farm surveys is that health authorities and the general public should continue to regard all cattle farms as potentially harbouring *E. coli* O157 and continue to recognise the significance of cattle as a potential source of human infection. To mitigate the risk to humans from contact with cattle and the farm environment, current advice regarding hygiene measures should be maintained.

Based on practices discussed in Sweden, potentially further advice and information could be provided to the public and other stakeholders in areas with higher densities of ruminants and *E. coli* O157 types that are a serious threat to human health. There is value here to use generic messages as pathogenic *E. coli* are only one of a number of zoonotic infections that can be acquired from these environments. We recommend holding a stakeholder meeting to discuss the main findings of this report and its practical applications. We also propose that a specific video documentary be made about the history and current issues around *E. coli* O157 in Scotland.

5.1.2 On farm control including vaccine field testing

Our research validates the use of vaccination to reduce *E. coli* O157 excretion from cattle. This has been demonstrated to reduce animal-to-animal transmission under controlled laboratory conditions and, in turn, modelling indicates this will lower the likelihood of human infection from cattle sources. The key next step towards uptake of vaccination as an intervention is evaluation under field conditions. This will require support for vaccine safety trials as trials on farms cannot realistically be conducted until a vaccine is under licence. The commercial drivers for such a vaccine and who should support the costs are a serious issue for any further development. Therefore these next steps will require support from a public health and political agenda that recognises that commercial interests in a vaccine are tempered by lack of efficacy of previous formulations and the fact that drivers for its use are separate from animal health concerns. This can be supported by additional modelling work exploring the potential impact of the vaccine at a GB-wide rather than a Scottish level.

Such a recommendation is made in the context that other specific interventions may also be valid including pre/probiotics, phage treatment, dietary manipulation and general biosecurity.

5.1.3 Continued implementation of whole genome sequencing

The precision of WGS within outbreak investigations provides more information on the likely source of an isolate, not just whether it is a local or imported strain but potentially the region in the UK it may originate from. Such information can provide information on the vehicles of transmission and processes that lead to human infection. As the sequence and associated metadata collections increase, then the

‘mapping’ accuracy of such data will also improve and therefore we should continue to collect and sequence isolates from animal and environmental sources to compare these with human isolates.

WGS can also provide insights into the relative virulence of an isolate and again the more information that can be provided on the disease and epidemiology around any isolate sequence, the more can be predicted from any new isolate sequence. This has to work within ethical constraints of data use, but we recommend that going forward, non-sensitive meta-data records are generated by agencies such as HPS and PHE that can be made available when associated sequences are released into public databases. There is a cost element to this, but if Scotland and the UK aim to be a renowned centre for ‘One Health’ activities then this type of data access is critical. We also recommend that long read sequencing technologies should be used alongside short-read approaches in significant outbreaks as we have shown that such data can add further information about the provenance of an isolate and highlight emerging threats.

5.2. Future Work

5.2.1. Mapping of human and animal isolates.

As initially anticipated in this project, there is a need for location and exposure data for human infections which then can be related with regional farm isolates using the resolution of sequence data. Specifically, we should then be able to confirm whether the higher rate of *E. coli* O157 human infections in Scotland is due to higher levels of local exposure to specific subtypes of *stx*2a+ *E. coli* O157 compared to that occurring in England & Wales.

5.2.2 Further research on the role of Shiga toxins in ruminant and environmental reservoirs.

Further research is needed to understand fundamentally how Shiga toxin provides an advantage for colonisation and transmission of *E. coli* in cattle. This understanding could lead to approaches to reduce selection of this toxin in the ruminant reservoir, for example through feed-based approaches.

5.2.3 National surveys of sheep flocks and dairy herds.

There is a lack of knowledge about the strains of STEC present in the dairy sector and in other ruminants such as sheep. We recommend that surveys are conducted of these other potential reservoirs to determine prevalence and circulating strain types. This will help us to understand the relative risks from such hosts. It would augment ongoing work investigating the epidemiology of *E. coli* O157 in deer in Scotland.

5.2.4 Further research to understand vehicles of transmission including development of food testing methods.

Unprocessed or minimally processed foods can, very rarely, be contaminated with *E. coli* O157 and other STEC serotypes that pose a threat to human health. General screening is not feasible and the main protection is hazard control and prevention of cross contamination in the production process, as well as guidance on how foods may be treated before consumption. However, when outbreaks do occur there is still a need to screen large numbers of food and other samples for STEC and suitable methods still need to be developed to facilitate this process as current PCR and culture methods are expensive, time consuming and labour intensive. Such methods could then be applied in screening programmes to learn more about other routes of transmission and exposure levels.

5.2.5 Vaccine feedtrials

In line with one of our recommendations we are trying to obtain funding to support a feedlot trial of the vaccine formulation in the USA. This would test the vaccine with much larger animal numbers and in a natural production setting. Such a trial is required to provide investors with the confidence to support larger-scale production and safety tests. The impact of the vaccine on the excretion of key non-O157 STEC serotypes should also be investigated.

5.2.6 Risk factor analyses

The geographical distribution of *E. coli* O157 positive farms from all 3 surveys (SEERAD (1998-2000), IPRAVE (2002-2004), BECS (2014-2015)) could be combined with data on farm demography, management, water source, animal movement and various environmental variables to look for areas within Scotland that are at increased risk of *E. coli* O157.

5.2.7. Added value from projects.

Escherichia coli is a common bacterial species that can be isolated from many types of animal and environmental samples. As such it has and continues to be used as a sentinel organism, initially to provide evidence of faecal contamination such as in water courses but more recently as an indicator of antimicrobial resistance. Funding bodies need to consider added value from their surveys and ensure that archived material and strain collections are made available to optimise outputs to the community.

5.3. Knowledge Exchange

Output Publications (OP)

Published

OP1 Henry MK, Tongue SC, Evans J, Webster C, McKendrick IJ, Morgan M, Willett A, Reeves A, Humphry RW, Gally DL, Gunn GJ, Chase-Topping ME (2017). British *Escherichia coli* O157 in Cattle Study (BECS): to determine the prevalence of *E. coli* O157 in herds with cattle destined for the food chain. *Epidemiol Infect*. 2017 145:3168-3179 doi: 10.1017/S0950268817002151

OP2 Dallman TJ, Ashton PM, Byrne L, Perry NT, Petrovska L, Ellis R, Allison L, Hanson M, Holmes A, Gunn GJ, Chase-Topping ME, Woolhouse ME, Grant KA, Gally DL, Wain J, Jenkins C (2015). 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. doi: 10.1099/mgen.0.000029.

OP3 Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL (2016). Support vector machine applied to predict the zoonotic potential of *E. coli* O157 cattle isolates. *Proc Natl Acad Sci USA*. 2016 113:11312-11317.

OP4 Holmes A, Dallman T, Shabaan S, Hanson M, Allison L (2018). Validation of Whole Genome Sequencing for Identification and Characterisation of Shiga Toxin-Producing *Escherichia coli* to produce standardised data to enable data sharing. Journal of Clinical Microbiology. *J. Clin. Microbiol.* Accepted manuscript posted online *20 December 2017*, doi:*10.1128/JCM.01388-17*

OP5 Shaaban S, Cowley LA, McAteer SP, Jenkins C, Dallman TJ, Bono JL, Gally DL (2016). Evolution of a zoonotic pathogen: investigating prophage diversity in enterohaemorrhagic *Escherichia coli* O157 by long-read sequencing. *Microb Genom*. 2016 2(12):e000096. doi: 10.1099/mgen.0.000096.

OP6 Cowley LA, Dallman TJ, Fitzgerald S, Irvine N, Rooney PJ, McAteer SP, Day M, Perry NT, Bono JL, Jenkins C, Gally DL (2016). Short-term evolution of Shiga toxin-producing Escherichia coli O157:H7 between two food-borne outbreaks. *Microb Genom*. 2016 2(9):e000084. doi: 10.1099/mgen.0.000084.

OP7 Schutz K, Cowley LA, Shaaban S, Carroll A, McNamara E, Gally DL, Godbole G, Jenkins C, Dallman TJ. Evolutionary Context of Non-Sorbitol-Fermenting Shiga Toxin-Producing *Escherichia coli* O55:H7. *Emerg Infect Dis*. 2017 23:1966-1973. doi: 10.3201/eid2312.170628.

OP8 Howell AK, Tongue SC, Currie C, Evans J, Williams DJL, McNeilly TN. (2018) Co-infection with Fasciola hepatica may increase the risk of Escherichia coli O157 shedding in British cattle destined for the food chain. Howell AK, Tongue SC, Currie C, Evans J, Williams DJL, McNeilly TN. *Prev Vet Med*. 1;150:70-76. doi: 10.1016/j.prevetmed.2017.12.007.

OP9 Prentice JC, Marion G, Hutchings MR, McNeilly TN, Matthews L. (2017) Complex responses to movement-based disease control: when livestock trading helps. *J R Soc Interface.* 14(126). pii: 20160531. doi: 10.1098/rsif.2016.0531.

OP10 Hickey GL, Diggle PJ, McNeilly TN, Tongue SC, Chase-Topping ME, Williams DJ. (2015) The feasibility of testing whether *Fasciola hepatica* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol. *Prev Vet Med.* 119(3-4):97-104. doi: 10.1016/j.prevetmed.2015.02.022.

In preparation

1. Fitzgerald S, Beckett AE, Palarea-Albaladejo J, McAteer S, Shaaban S, Morgan J, Ahmad NI, Bono JL, Gally DL, McNeilly TN. Shiga toxin sub-type 2a increases the efficiency of *Escherichia coli* O157 transmission between animals. *In preparation for PloS Pathogens*

2. Beckett AE, Morgan J, Corbishley AC, Corripio-Miyar Y, Fitzgerald S, Frew D, McAteer SP, Gally DL, and McNeilly TN. Defining bystander effects of colonisation with Shigatoxigenic *Escherichia coli* on adaptive immune responses in cattle *In preparation for Infection and Immunity*

3. Immunization with a Shiga toxoid-based vaccine results in enhanced humoral immunity in cattle naturally exposed to Shiga toxigenic *Escherichia coli.* Beckett AE, Schmidt N, Barth SA, Gally DL, Bono JL, Rong W, Geue L, Menge C, McNeilly TN. *In preparation for Veterinary Research*.

4. The British *Escherichia coli* O157 in Cattle Study (BECS): Factors associated with the occurrence of *E. coli* O157 Henry MK, McCann C, Humphry RW, Evans J, Webster C, Morgan M, Willett A, Gunn GJ, Chase-Topping ME, Tongue SC. *In preparation for BMC Veterinary Research*.

Scientific conferences:

VTEC 2015 9th International Symposium. Consortium members had a substantial presence at this international meeting; including: D. Gally invited speaker; T. Dallman invited speaker and multiple poster presentations including: (1) Defining bystander effects of colonisation with Shigatoxigenic Enterohaemorrhagic *Escherichia coli* O157 on adaptive immune responses in cattle (Poster presentation); (2) Cellular immunity to Enterohaemorrhagic *Escherichia coli* O157 in cattle and implications for vaccine development (Poster presentation).

World Buiatrics Congress July 2016, Dublin Ireland: Abstract in conference proceedings and oral presentation videoed for conference web presentation – “The prevalence of verocytotoxigenic *E. coli* O157 (VTEC O157) in British cattle populations” M K Henry.

Society for Veterinary Epidemiology and Preventive Medicine (SVEPM) Annual Conference March 2017. J. Prentice presented the meta population modelling work.

International Association of Food Protection Annual Conference, Florida, USA, July 2017: D. Gally invited speaker; ‘What Can We Do with 10,000 Genomes That Couldn't be Done with 100?’.

Med-Vet-Net Association 5th International Scientific Conference, University of Surrey, D. Gally invited speaker. 'STEC O157 transmission biology and vaccine development' July 2017.

Institute of Biomedical Science (IBMS) Congress 2017, Birmingham (Sept 2017): Invited oral presentation by J. Evans "VTEC: More to the story than cows and people".

FSIS, Washington, USA: Public meeting - Use of Whole Genome Sequence (WGS) Analysis to Improve Food Safety and Public Health. D. Gally invited speaker: setting the scene Advanced technology (including WGS) to improve animal health and food quality. Oct. 2017.

European College Veterinary Public Health AGM and Scientific Conference October 2017, Liege, Belgium - Abstract, short oral presentation and poster; poster on-line ECVPH conference site - The prevalence of verocytotoxigenic *E. coli* O157 (VTEC O157) in British cattle populations.

Super-shedding International Workshop, Edinburgh. Nov. 2017 supported by the FSS/FSA super-shedding programme and a BBSRC international partnership award.

SVEPM March 2018, Tallin, Estonia. Presentation of the meta population modelling work. Poster (on-line on website) The prevalence of verocytotoxigenic *E. coli* O157 (VTEC O157) in British cattle populations; S. Tongue.

Microbiology Society, Birmingham, April 2018. Gally invited speaker. The different scales of microevolution in enterohaemorrhagic *Escherichia coli* O157 and relationship with zoonotic threat.

VTEC 2018 10th international conference in Florence. Again the consortium had a significant presence at the conference including invited speakers as well as selected oral and poster presentations:

Abstract submitted April 2018 for International Society of Veterinary Epidemiology and Economics (November 2018) – The British *E. coli* in O157 study (BECS): more than just prevalence estimates – implications for public health. -– poster presentation accepted S.Tongue

Abstract submitted for a poster presentation for British Cattle Veterinary Association Congress (October 2018) - The British *Escherichia coli* O157 in cattle study (BECS): prevalence and factors associated with occurrence in British cattle populations - accepted but withdrawn at the last minute due to unforeseen circumstances – to be re-submitted with more of a knowledge exchange, stakeholder messages angle, as either oral or poster for British Cattle Veterinary Association Congress (October 2019)

***National press**:*

Press release: "Volunteer farmers sought for nationwide dung sampling project"

- (August 2014) SRUC News Release No 14R133 by the Epidemiology Research Unit and SRUC Communications Team led to

- BBC Radio Scotland News Drive interview (G J Gunn) with Bill Whitehead

Vaccine developed to fight *E. coli* (The Press and Journal, 17/12/2016): https://www.pressandjournal.co.uk/fp/business/farming/1115539/vaccine-developed-to-fight-e-coli/

New vaccine to fight *E. coli* developed in Scotland (The Courier & Advertiser, 14/12/2016): https://www.thecourier.co.uk/fp/business/farming/farming-news/327717/new-vaccine-fight-e-coli-developed-scotland/

Researchers in Scotland develop *E. coli* vaccine for cattle (Food Safety News, 21/12/2016, http://www.foodsafetynews.com/2016/12/researchers-in-scotland-develop-e-coli-vaccine-for-cattle/#.WpPmdnynwdU)

'The computer says food poisoning' The Times page 16, 21/09/16

STV evening news: 20/09/16 about machine learning prediction of virulent E. coli O157 (video no longer available)

Computers learn to spot deadly bacteria: Science Daily: https://www.sciencedaily.com/releases/2016/09/160921163340.htm

Also Daily Express:

https://www.express.co.uk/life-style/health/712541/E-coli-food-poisoning-bacteria-computer-detection

Dangerous *E coli* strain 'is evolving' The Times page 10, 23/09/15

E. coli bug became potentially deadly in the 1980's. The Herald, page 11, 23/09/15

Deadly toxin turned *E.coli* into killer disease. The Scotsman, page 10, 23/09/15

'Health warnings on unpasteurised cheese 'should be considered'' Herald 29/09/17: http://www.heraldscotland.com/news/health/15565022.Health_warnings_on_unpasteurised_cheese__should_be_considered_/

Government and stakeholders

Letters to stakeholders and industry at start of field survey (August 2014) *E. coli* prevalence study among finishing cattle in the UK, Henry M; Tongue S; Gunn G in Veterinary Record, 2014:175:208 doi 10.1136/vr.g5328 Veterinary Times, SAC Consulting FRBS & Veterinary Services and APHA.

Included in, or the main subject of, at least 20 oral presentations by SRUC project team members to external parties at meetings, workshops or seminars including agricultural consultants, practicing vets and scientific researchers in other fields.

Policy brief provided on current *E. coli* O157 research in Scotland for the Animal Health and Welfare Division of the Scottish Government (December 2017).

Parliamentary motion S5M-01599 (21/09/16) University of Edinburgh Research into Use of Software to tackle Food Poisoning.

Feedback to participating farms – APPENDIX G

5.4. Personnel & Training

PhD studentship 1: AB has now completed her trial and laboratory work and is currently writing up her Ph.D. Amy has obtained a Veterinary Field Officer position on the Isle of Man. Two main publications are being drafted from her work.

PhD studentship 2: SS has now completed his work and is currently writing up his PhD. He is now working for Scottish Microbiology Reference Laboratories at NHS Greater Glasgow & Clyde as a Bioinformatician. He has contributed to two publications, one as primary author.

JS started the programme as the main PDRA working on shedding from cattle but left for a permanent position with FDA after 6 months.

SF was the main PDRA focused on excretion experiments and analysis of *E. coli* O157 variation *in vivo* and *in vitro*. He has obtained more funding (BBSRC) to continue *E. coli* O157 research as a PDRA at Edinburgh.

JP was a PDRA focused on statistical modelling of the vaccine trials and metapopulation modelling of the potential effectiveness of the vaccine in the Scottish cattle population. He has obtained more PDRA funding to continue to work on the epidemiology of livestock disease at the University of Glasgow.

AT was a Masters student seconded at University of Glasgow and gained experience in epidemiological modelling.

MH is a field veterinary epidemiologist, ERU SRUC and gained experience in the development and management required - as well as the epidemiological aspects - of running a large scale collaborative field study (BECS), in addition to quantitative analysis and production of a variety of outputs, for different audiences.

JE and CW gained experience in the use of RT-PCR, adding this facility to the ERU lab capabilities; with scope for use in other areas of investigation.

5.5. Added Value Projects

BECS added-value projects

Projects that SRUC are involved in contributing to, or collaborating in, that have been leveraged or made possible by the existence of the BECS field survey samples include:

- Exploring whether there is an association between co-infection with VTEC O157 and liver fluke - Dr Tom McNeilly (Moredun Research Institute) and Prof. Diana Williams, University of Liverpool
  Outputs to date 14/12/17:
  - OP8: K. Howell, **S. C. Tongue**, C. Currie, **J. Evans**, D. J. L. Williams, **T. N. McNeilly** Co-infection with Fasciola hepatica may increase the risk of *Escherichia coli* O157 shedding in British cattle destined for the

food chain. Preventive Veterinary Medicine - accepted 06/12/17 doi: to be advised

  - OP10: G.L. Hickey, P.J. Diggle, **T.N. McNeilly, S.C. Tongue**, M.E. Chase-Topping, D.J.L. Williams. The feasibility of testing whether *Fasciola hepatica* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol. Preventive Veterinary Medicine (2015) 119 (3-4) 97-104 http://dx.doi.org/10.1016/j.prevetmed.2015.02.022

  - Poster and abstract included in conference proceedings for the paper above entitled “Is it feasible to test if liver fluke (*Fasciola hepatica)* is associated with increased risk of verocytotoxin producing *Escherichia coli* O157 from an existing study protocol?” for two conferences: ISVEE 14 in Mexico 2015 and WAAVP 2015 in Liverpool.

- VM0526 – a survey for and characterisation of antibiotic resistance in Enterobacteriaceae from beef cattle in England & Wales and Scotland, with particular emphasis on Extended Spectrum Beta-Lactamase (ESBL) and carbapenemase resistance - Animal and Plant Health Agency (APHA). Outputs to date 14/12/17:
  - Poster and abstract included in conference proceedings entitled “Extended Spectrum Beta-Lactamase (ESBL) resistance in Enterobacteriaceae in beef cattle in Great Britain in 2015” for European Congress of Clinical Microbiology and Infectious Diseases Meeting, April 2017, Vienna Austria
  - Paper – submitted to Journal of Applied Microbiology and Final report with funders; not yet published

- Investigation of aspects of non-O157 STECs - (Wellcome Trust Career Re-entry Fellow at The Roslin Institute, University of Edinburgh)

6. Bibliography

1. Pennington H. The Public Inquiry into the September 2005 Outbreak of *E. coli* O157 in South Wales. http://gov.wales/docs/dhss/publications/150618ecoli-reporten.pdf.

2. Food Standards Agency. Understanding of the Factors That Lead to EHEC Colonisation in Cattle and the Role of Supershedding in the Transmission and Maintenance of Infection. 2011.

3. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet. 2005. doi:10.1016/S0140-6736(05)71144-2.

4. Zhang Y, Laing C, Steele M, et al. Genome evolution in major *Escherichia coli* O157:H7 lineages. BMC Genomics. 2007;8:121. doi:10.1186/1471-2164-8-121.

5. Manning SD, Motiwala AS, Springman AC, et al. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc Natl Acad Sci. USA. 2008. doi:10.1073/pnas.0710834105.

6. Naylor SW, Low JC, Besser TE, et al. Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host. Infect Immun. 2003;71(3). doi:10.1128/IAI.71.3.1505-1512.2003.

7. Low JC, McKendrick IJ, McKechnie C, et al. Rectal carriage of enterohemorrhagic *Escherichia coli* O157 in slaughtered cattle. Appl Environ Microbiol. 2005;71(1):93-97. doi:10.1128/AEM.71.1.93-97.2005.

8. Sheng H, Davis MA, Knecht HJ, Hovde CJ. Rectal administration of *Escherichia coli* O157:H7: Novel model for colonization of ruminants. Appl Environ Microbiol. 2004. doi:10.1128/AEM.70.8.4588-4595.2004.

9. Naylor SW, Nart P, Sales J, Flockhart A, Gally DL, Low JC. Impact of the direct application of therapeutic agents to the terminal recta of experimentally colonized calves on *Escherichia coli* O157:H7 shedding. Appl Environ Microbiol. 2007;73(5). doi:10.1128/AEM.01736-06.

10. Naylor SW, Roe AJ, Nart P, et al. *Escherichia coli* O157:H7 forms attaching and effacing lesions at the terminal rectum of cattle and colonization requires the LEE4 operon. Microbiology. 2005;151:2773-2781. doi:10.1099/mic.0.28060-0.

11. Xu X, McAteer SP, Tree JJ, et al. Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli*. PLoS Pathog. 2012;8(5):e1002672. doi:10.1371/journal.ppat.1002672.

12. Dziva F, van Diemen PM, Stevens MP, Smith AJ, Wallis TS. Identification of Escherichia coli O157:H7 genes influencing colonization of the bovine

gastrointestinal tract using signature-tagged mutagenesis. Microbiology. 2004. doi:10.1099/mic.0.27448-0.

13. Wong ARC, Pearson JS, Bright MD, et al. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: Even more subversive elements. Mol Microbiol. 2011. doi:10.1111/j.1365-2958.2011.07661.x.

14. Tobe T, Beatson SA, Taniguchi H, et al. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc Natl Acad Sci USA. 2006;103(40):14941-14946. doi:10.1073/pnas.0604891103.

15. Dean P, Kenny B. The effector repertoire of enteropathogenic *E. coli*: ganging up on the host cell. Curr Opin Microbiol. 2009. doi:10.1016/j.mib.2008.11.006.

16. Tree JJ, Wolfson EB, Wang D, Roe AJ, Gally DL. Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. Trends Microbiol. 2009;17(8):361-370. doi:10.1016/j.tim.2009.06.001.

17. Tree JJ, Roe AJ, Flockhart A, et al. Transcriptional regulators of the GAD acid stress island are carried by effector protein-encoding prophages and indirectly control type III secretion in enterohemorrhagic *Escherichia coli* O157:H7. Mol Microbiol. 2011;80(5). doi:10.1111/j.1365-2958.2011.07650.x.

18. Tildesley MJ, Gally DL, McNeilly TN, Low JC, Mahajan A, Savill NJ. Insights into mucosal innate responses to *Escherichia coli* O157 : H7 colonization of cattle by mathematical modelling of excretion dynamics. J R Soc Interface. 2012;9(68):518-527. doi:10.1098/rsif.2011.0293.

19. Menge C, Blessenohl M, Eisenberg T, Stamm I, Baljer G. Bovine Ileal Intraepithelial Lymphocytes Represent Target Cells for Shiga Toxin 1 from Escherichia coli. Infect Immun. 2004. doi:10.1128/IAI.72.4.1896-1905.2004.

20. Menge C, Wieler LH, Schlapp T, Baljer G. Shiga toxin 1 from *Escherichia coli* blocks activation and proliferation of bovine lymphocyte subpopulations in vitro. Infect Immun. 1999, 67(5): 2209–2217.

21. Gobert AP, Vareille M, Glasser A-L, Hindre T, de Sablet T, Martin C. Shiga Toxin Produced by Enterohemorrhagic *Escherichia coli* Inhibits PI3K/NF- B Signaling Pathway in Globotriaosylceramide-3-Negative Human Intestinal Epithelial Cells. J Immunol. 2007. doi:10.4049/jimmunol.178.12.8168.

22. Robinson CM, Sinclair JF, Smith MJ, O'Brien AD. Shiga toxin of enterohemorrhagic *Escherichia coli* type O157:H7 promotes intestinal colonization. Proc Natl Acad Sci USA. 2006. doi:0602359103 [pii]\r10.1073/pnas.0602359103.

23. Sinclair JF, O'Brien AD. Cell surface-localized nucleolin is a eukaryotic receptor for the adhesin intimin-γ of enterohemorrhagic *Escherichia coli* O157:H7. J Biol Chem. 2002. doi:10.1074/jbc.M110230200.

24. Eckert SE, Dziva F, Chaudhuri RR, et al. Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. J Bacteriol. 2011. doi:10.1128/JB.01292-10.

25. Meltz Steinberg K, Levin BR. Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. Proc R Soc B Biol Sci. 2007. doi:10.1098/rspb.2007.0245.

26. Ravva S V., Sarreal CZ, Mandrell RE. Altered Protozoan and Bacterial Communities and Survival of *Escherichia coli* O157:H7 in Monensin-Treated Wastewater from a Dairy Lagoon. PLoS One. 2013. doi:10.1371/journal.pone.0054782.

27. WHO Report 2018: MICROBIOLOGICAL RISK ASSESSMENT SERIES 31 REPORT Shiga toxin-producing *Escherichia coli* (STEC) and food: attribution, characterisation, and monitoring. http://apps.who.int/iris/bitstream/handle/10665/272871/9789241514279-eng.pdf?sequence=1&isAllowed=y.

28. Beutin L, Geier D, Steinrück H, Zimmermann S, Scheutz F. Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. J Clin Microbiol. 1993;31(9):2483-2488. http://www.ncbi.nlm.nih.gov/pubmed/8408571.

29. Wieler LH, Vieler E, Erpenstein C, et al. Shiga toxin-producing *Escherichia coli* strains from bovines: association of adhesion with carriage of eae and other genes. J Clin Microbiol. 1996;34(12):2980-2984. http://www.ncbi.nlm.nih.gov/pubmed/8940434.

30. Chase-Topping M, Gally D, Low C, Matthews L, Woolhouse M. Super-shedding and the link between human infection and livestock carriage of *Escherichia coli* O157. Nat Rev Microbiol. 2008;6(12):904-912. doi:10.1038/nrmicro2029.

31. Spencer SEF, Besser TE, Cobbold RN, French NP. "Super" or just "above average"? Supershedders and the transmission of *Escherichia coli* O157:H7 among feedlot cattle. J R Soc Interface. 2015;12(110). doi:10.1098/rsif.2015.0446.

32. Wang O, McAllister TA, Plastow G, Stanford K, Selinger B, Guan LL. Interactions of the Hindgut Mucosa-Associated Microbiome with Its Host Regulate Shedding of *Escherichia coli* O157:H7 by Cattle. Björkroth J, ed. Appl Environ Microbiol. 2018;84(1). doi:10.1128/AEM.01738-17.

33. Zaheer R, Dugat-Bony E, Holman D, et al. Changes in bacterial community composition of *Escherichia coli* O157:H7 super-shedder cattle occur in the lower intestine. Smidt H, ed. PLoS One. 2017;12(1):e0170050. doi:10.1371/journal.pone.0170050.

34. Jacob ME, Paddock ZD, Renter DG, Lechtenberg KF, Nagaraja TG. Inclusion of Dried or Wet Distillers' Grains at Different Levels in Diets of Feedlot Cattle Affects Fecal Shedding of *Escherichia coli* O157:H7. Appl Environ Microbiol. 2010;76(21):7238-7242. doi:10.1128/AEM.01221-10.

35. Paddock ZD, Renter DG, Shi X, Krehbiel CR, DeBey B, Nagaraja TG. Effects of feeding dried distillers grains with supplemental starch on fecal shedding of *Escherichia coli* O157:H7 in experimentally inoculated steers. J Anim Sci. 2013;91(3):1362-1370. doi:10.2527/jas.2012-5618.

36. Wang O, McAllister TA, Plastow G, Stanford K, Selinger B, Guan LL. Host mechanisms involved in cattle *Escherichia coli* O157 shedding: a fundamental understanding for reducing foodborne pathogen in food animal production. Sci Rep. 2017;7(1):7630. doi:10.1038/s41598-017-06737-4.

37. Wang O, Liang G, McAllister TA, et al. Comparative Transcriptomic Analysis of Rectal Tissue from Beef Steers Revealed Reduced Host Immunity in *Escherichia coli* O157:H7 Super-Shedders. Kobeissy FH, ed. PLoS One. 2016;11(3):e0151284. doi:10.1371/journal.pone.0151284.

38. Matthews L, Low JC, Gally DL, et al. Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. Proc Natl Acad Sci USA. 2006;103(3):547-552. doi:10.1073/pnas.0503776103.

39. Chase-Topping ME, McKendrick IJ, Pearce MC, et al. Risk factors for the presence of high-level shedders of *Escherichia coli* O157 on Scottish farms. J Clin Microbiol. 2007. doi:10.1128/JCM.01690-06.

40. Pearce MC, Chase-Topping ME, McKendrick IJ, et al. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998-2000 and 2002-2004. BMC Microbiol. 2009. doi:10.1186/1471-2180-9-276.

41. Matthews L, Reeve R, Gally DL, et al. Predicting the public health benefit of vaccinating cattle against *Escherichia coli* O157. Proc Natl Acad Sci USA. 2013;110(40):16265-16270. doi:10.1073/pnas.1304978110.

42. Kim J, Nietfeldt J, Benson AK. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc Natl Acad Sci. USA 1999. doi:10.1073/pnas.96.23.13288.

43. Yang Z, Kovar J, Kim J, et al. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. Appl Environ Microbiol. 2004. doi:10.1128/AEM.70.11.6846-6854.2004.

44. Eppinger M, Mammel MK, Leclerc JE, Ravel J, Cebula TA. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. Proc Natl Acad Sci. USA 2011. doi:10.1073/pnas.1107176108.

45. Johnson RP, Gyles CL, Huff WE, et al. Bacteriophages for prophylaxis and therapy in cattle, poultry and pigs. Anim Heal Res Rev. 2008. doi:10.1017/S1466252308001576.

46. Snedeker KG, Campbell M, Sargeant JM. A Systematic Review of Vaccinations to Reduce the Shedding of *Escherichia coli* O157 in the Faeces of Domestic Ruminants. Zoonoses Public Health. 2012. doi:10.1111/j.1863-2378.2011.01426.x.

47. Sargeant JM, Amezcua MR, Rajic A, Waddell L. Pre-harvest interventions to reduce the shedding of *E. coli* O157 in the faeces of weaned domestic ruminants: A systematic review. Zoonoses Public Health. 2007. doi:10.1111/j.1863-2378.2007.01059.x.

48. Zhao T, Doyle MP, Harmon BG, Brown CA, Mueller POE, Parks AH. Reduction of carriage of enterohemorrhagic *Escherichia coli* O157:H7 in cattle by inoculation with probiotic bacteria. J Clin Microbiol. 1998.

49. FSA. Feasibility of Introducing Methods, in the UK, for Reducing Shedding of *E. coli* O157 in Cattle.; 2013.

50. Wisener L V., Sargeant JM, O'Connor AM, Faires MC, Glass-Kaastra SK. The Use of Direct-Fed Microbials to Reduce Shedding of *Escherichia coli* O157 in Beef Cattle: A Systematic Review and Meta-analysis. Zoonoses Public Health. 2015;62(2):75-89. doi:10.1111/zph.12112.

51. McNeilly TN, Mitchell MC, Rosser T, et al. Immunization of cattle with a combination of purified intimin-531, EspA and Tir significantly reduces shedding of *Escherichia coli* O157:H7 following oral challenge. Vaccine. 2010;28(5):1422-1428. doi:10.1016/j.vaccine.2009.10.076.

52. McNeilly TN, Mitchell MC, Nisbet AJ, et al. IgA and IgG antibody responses following systemic immunization of cattle with native H7 flagellin differ in epitope recognition and capacity to neutralise TLR5 signalling. Vaccine. 2010;28(5). doi:10.1016/j.vaccine.2009.10.148.

53. McNeilly TN, Naylor SW, Mahajan A, et al. *Escherichia coli* O157:H7 colonization in cattle following systemic and mucosal immunization with purified H7 flagellin. Infect Immun. 2008;76(6). doi:10.1128/IAI.01452-07.

54. McNeilly TN, Mitchell MC, Corbishley A, et al. Optimizing the protection of Cattle against *Escherichia coli* O157:H7 colonization through immunization with different combinations of H7 flagellin, Tir, intimin-531 or EspA. PLoS One. 2015;10(5). doi:10.1371/journal.pone.0128391.

55. Thomson DU, Loneragan GH, Thornton AB, et al. Use of a Siderophore Receptor and Porin Proteins-Based Vaccine to Control the Burden of *Escherichia coli* O157:H7 in Feedlot Cattle. Foodborne Pathog Dis. 2009. doi:10.1089/fpd.2009.0290.

56. Thornton AB, Thomson DU, Loneragan GH, et al. Effects of a siderophore receptor and porin proteins-based vaccination on fecal shedding of *Escherichia coli* O157:H7 in experimentally inoculated cattle. J Food Prot. 2009, 72(4):866-9.

57. Henry MK, Tongue SC, Evans J, et al. British *Escherichia coli* O157 in Cattle Study (BECS): to determine the prevalence of *E. coli* O157 in herds with cattle destined for the food chain. Epidemiol Infect. 2017. doi:10.1017/S0950268817002151.

58. Gunn GJ, McKendrick IJ, Ternent HE, Thomson-Carter F, Foster G, Synge BA. An investigation of factors associated with the prevalence of verocytotoxin producing *Escherichia coli* O157 shedding in Scottish beef cattle. Vet J. 2007;174:554-564. doi:10.1016/j.tvjl.2007.08.024.

59. Pearce MC, Fenlon D, Low JC, et al. Distribution of *Escherichia coli* O157 in bovine fecal pats and its impact on estimates of the prevalence of fecal shedding. Appl Environ Microbiol. 2004. doi:10.1128/AEM.70.10.5737-5743.2004.

60. Evans J, Knight H, Mckendrick IJ, et al. Prevalence of *Escherichia coli* O157:H7 and serogroups O26, O103, O111 and O145 in sheep presented for slaughter in Scotland. J Med Microbiol. 2011. doi:10.1099/jmm.0.028415-0.

61. Tewolde R, Dallman T, Schaefer U, et al. MOST: a modified MLST typing tool based on short read sequencing. PeerJ. 2016. doi:10.7717/peerj.2308.

62. Joensen KG, Tetzschner AMM, Iguchi A, Aarestrup FM, Scheutz F. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol. 2015. doi:10.1128/JCM.00008-15.

63. Stamatakis A. Using RAxML to Infer Phylogenies. Curr Protoc Bioinformatics. 2015. doi:10.1002/0471250953.bi0614s51.

64. Holmes A, Jenkins C, Hanson M, et al. Applying phylogenomics to understand the emergence of Shiga Toxin producing *Escherichia coli* O157:H7 strains causing

severe human disease in the United Kingdom. Microb Genomics. 2015. doi:10.1099/mgen.0.000029.

65. Sinclair C, Jenkins C, Warburton F, Adak GK, Harris JP. Investigation of a national outbreak of STEC *Escherichia coli* O157 using online consumer panel control methods: Great Britain, October 2014. Epidemiol Infect. 2017;145(5):864-871. doi:10.1017/S0950268816003009.

66. Rowell S, King C, Jenkins C, et al. An outbreak of Shiga toxin-producing *Escherichia coli* serogroup O157 linked to a lamb-feeding event. Epidemiol Infect. 2016;144(12):2494-2500. doi:10.1017/S0950268816001229.

67. 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. doi:10.1017/S0950268816000509.

68. Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL. Support vector machine applied to predict the zoonotic potential of *E. coli* O157 cattle isolates. Proc Natl Acad Sci USA. 2016;113(40). doi:10.1073/pnas.1606567113.

69. Lupolova N, Dallman TJ, Holden NJ, Gally DL. Patchy promiscuity: machine learning applied to predict the host specificity of *Salmonella enterica* and *Escherichia coli*. Microb Genomics. 2017. doi:10.1099/mgen.0.000135.

70. Shaaban S, Cowley LA, McAteer SP, et al. Evolution of a zoonotic pathogen: investigating prophage diversity in enterohaemorrhagic *Escherichia coli* O157 by long-read sequencing. Microb Genomics. 2016. doi:10.1099/mgen.0.000096.

71. Schutz K, Cowley LA, Shaaban S, et al. Evolutionary context of non–sorbitol-fermenting Shiga toxin–producing *Escherichia coli* O55:H7. Emerg Infect Dis. 2017;23(12). doi:10.3201/eid2312.170628.

72. Cowley LA, Dallman TJ, Fitzgerald S, et al. Short term evolution of Shiga toxin producing *Escherichia coli* O157:H7 between two food-borne outbreaks. Microb Genomics. 2016. doi:10.1099/mgen.0.000084.

73. Ogura Y, Mondal SI, Islam MR, et al. The Shiga toxin 2 production level in enterohemorrhagic *Escherichia coli* O157:H7 is correlated with the subtypes of toxin-encoding phage. Sci Rep. 2015;5(1):16663. doi:10.1038/srep16663.

74. Hoffman MA, Menge C, Casey TA, Laegreid W, Bosworth BT, Dean-Nystrom EA. Bovine immune response to Shiga-toxigenic *Escherichia coli* O157:H7. Clin Vaccine Immunol. 2006;13(12):1322-1327. doi:10.1128/CVI.00205-06.

75. Schmidt CE, Shringi S, Besser TE. Protozoan Predation of *Escherichia coli* O157:H7 Is Unaffected by the Carriage of Shiga Toxin-Encoding Bacteriophages. DebRoy C, ed. PLoS One. 2016;11(1):e0147270. doi:10.1371/journal.pone.0147270.

76. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014. doi:10.1093/bioinformatics/btu170.

77. Ashton PM, Perry N, Ellis R, et al. Insight into Shiga toxin genes encoded by *Escherichia coli* O157 from whole genome sequencing. PeerJ. 2015. doi:10.7717/peerj.739.

78. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009. doi:10.1093/bioinformatics/btp324.

79. McKenna AH, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010. doi:10.1101/gr.107524.110.

80. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface. 2010. doi:10.1098/rsif.2009.0386.

APPENDICES

APPENDIX A: Sub-clustering and typing of *E. coli* O157: part 1

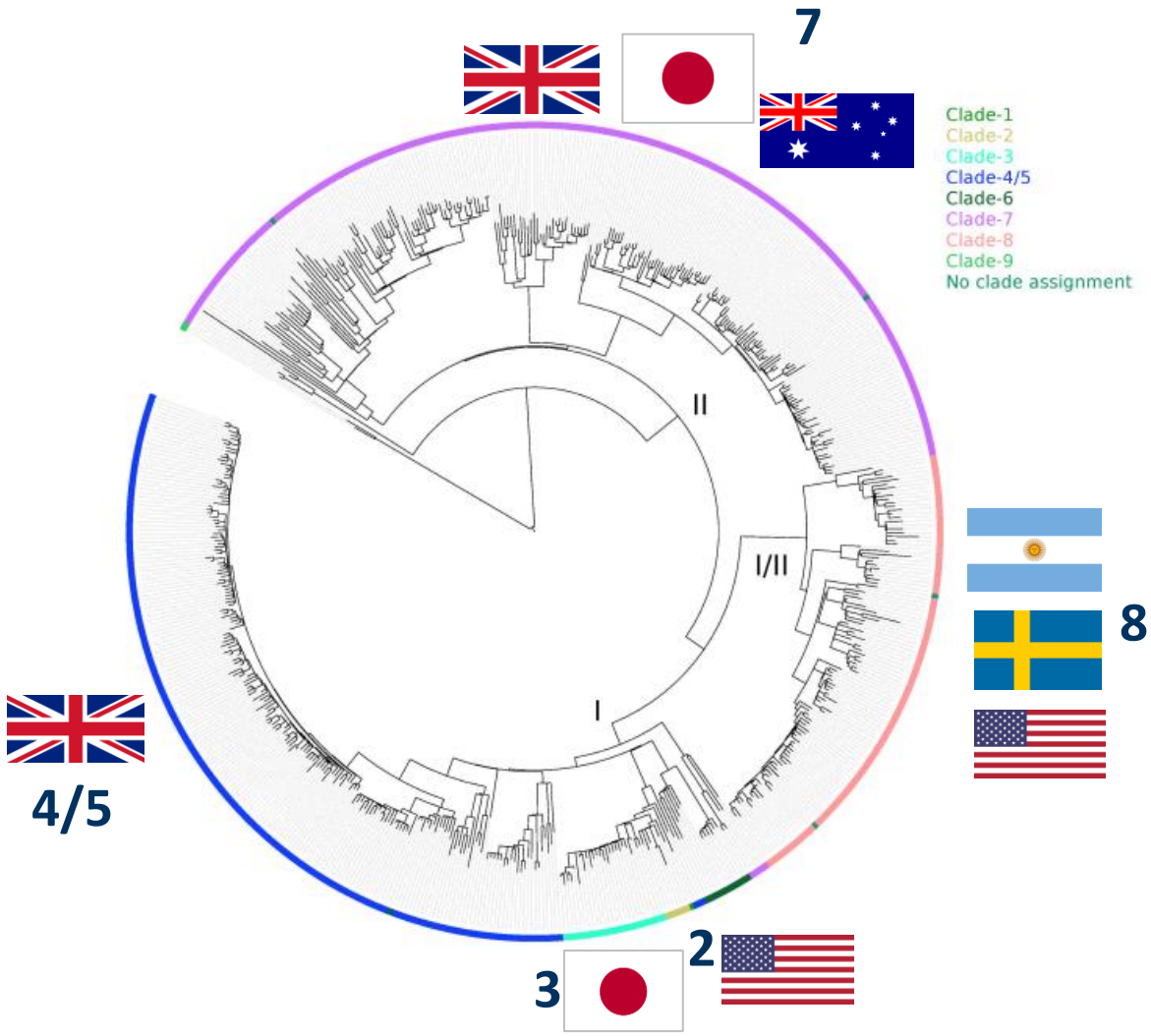
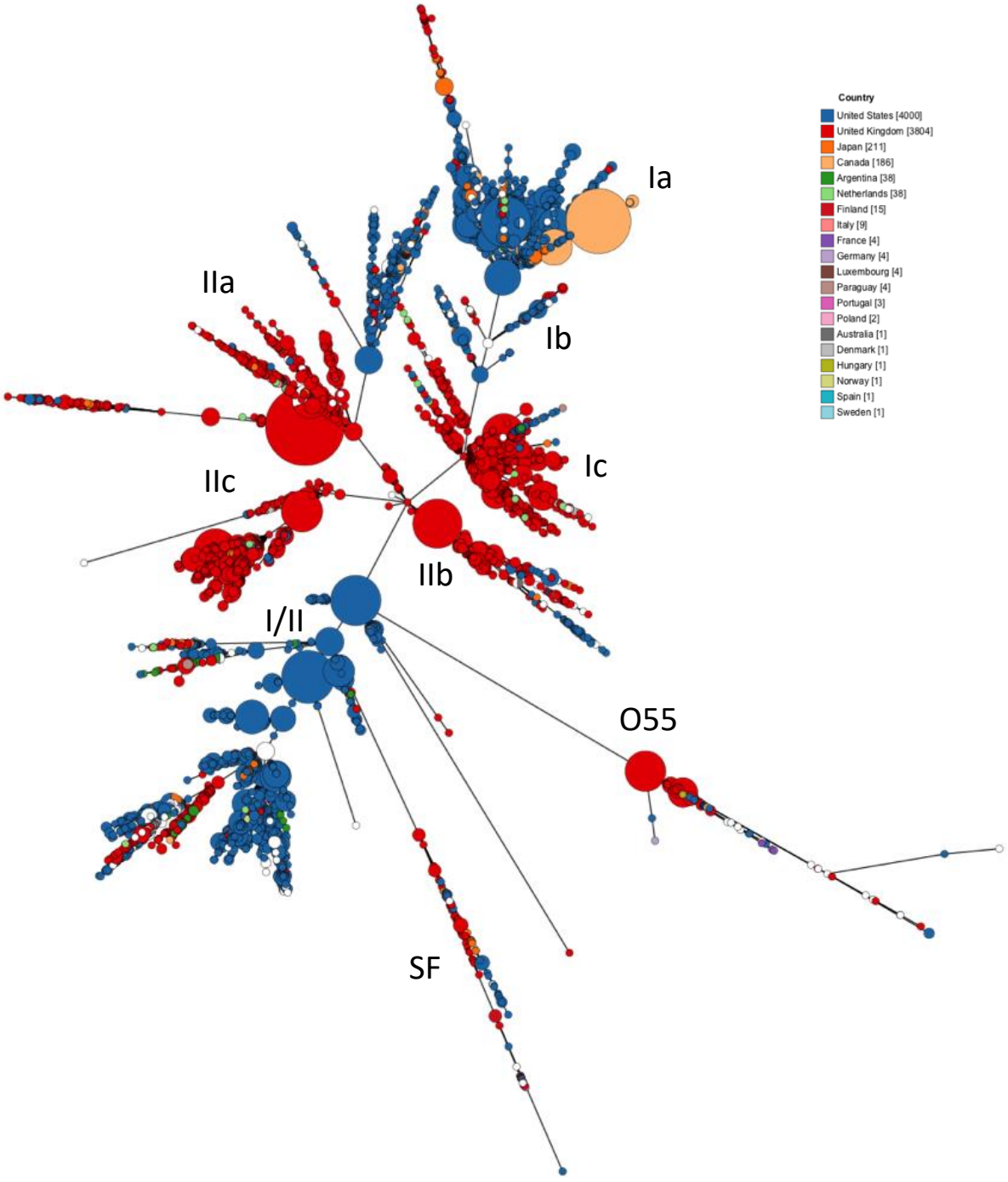


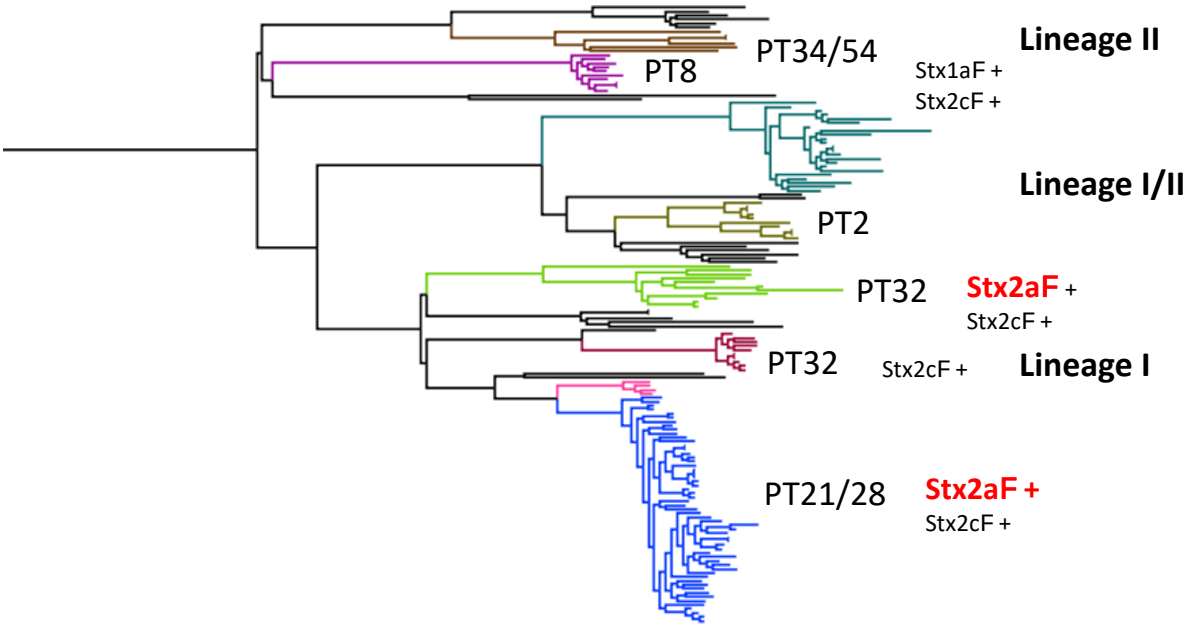
Figure based on data presented in Dallman *et al* 2015 Microbial Genomics 10.1099/mgen.0.000029 (OP2). It shows the clustering of ~1000 *E. coli* O157 sequences based on SNP relationship. The 3 main lineages are shown along with the 9 coloured clades. Country flags adjacent to key sub-types that predominate and are a human health issue. Our UK PT21/28 phage type is within Lineage 1, clade 4/5.

APPENDIX A: Sub-clustering and typing of *E. coli* O157: part 2



The figure shows relationships between *E. coli* O157 isolates and serotype O55 by country as defined by multi-locus sequence typing (MLST). A colour key for the countries is provided as well as an indication of overlap with the key lineage subtypes. SF = sorbitol fermenting.

APPENDIX A: Sub-clustering and typing of *E. coli* O157: part 3



The figure shows sub-clustering of *E. coli* O157 isolates (UK) based on WGS and then SNP relatedness to a reference strain. The main lineages are shown as are some key phage types and an indication of association with Shiga toxin types (a general rule, not strict as some vary). It can be seen there is a good association between the sequence-based clustering and phage type. Stx2a phage introduction coincides with emergence of O157 as a serious human infection and this occurred ~40 years ago. Please see Dallman *et al* 2015 Microbial Genomics 10.1099/mgen.0.000029 (OP2) for more details - hyperlink

APPENDIX B: Deliverables as Direct Objectives (DO) of the programme

	Deliverables		
	Please outline the proposed project outputs/deliverables. Please provide a timetable of key dates for the project (for example fieldwork dates, dates for provision of draft and final research materials, and reporting). Deliverables must be linked to the objectives. Please insert additional rows to the table below as required.		
	Organisation	Target Date	Deliverable details
1.1.1	MRI, UOE	01/10/2014	Excretion data for animals that have naturally-acquired a wild type PT21/28 (super-shedding lineage III) *E. coli* O157 strain and information relating excretion dynamics to shedding levels of the in-contact animals as well as levels of *E. coli* O157 in the local environment. This will include duration of excretion for up to 4 weeks.
1.1.2	MRI, UOE	01/07/2015	Excretion data of animals that have naturally-acquired a wild type PT32 (non super-shedding lineage IV) *E. coli* O157 strain and information relating excretion dynamics to shedding levels of the in-contact animals as well as levels of *E. coli* O157 in the local environment. This will include duration of excretion for up to 4 weeks
1.1.3	MRI, UOE	01/07/2015	Quantification of the extent of environmental contamination in the different groups.
1.1.4	MRI, UOE	01/09/2015	Measures of cellular and humoral responses of animals to key EHEC antigens before and during colonisation with two wild type *E. coli* O157 strains.
1.1.5	MRI, UOE	01/09/2015	Measures of innate responses from rectal pinch biopsies and both rectal follicles and local lymph nodes following isolation of these tissues at post

			mortem before and during colonisation with two wild type *E. coli* O157 strains.
1.1.6	MRI, BIOSS, UOE	01/12/2015	Statistical analyses of excretion data from PT21/28 and PT32 strains
1.1.7	MRI, UOE	01/05/2015	To provide numerical data for modelling under Objective 3.
1.2.1	MRI, UOE	01/10/2014	Transmission frequencies from animals that have naturally-acquired wild type PT21/28 *E. coli* O157 strains to in-contact naïve animals.
1.2.2	MRI, UOE	01/04/2015	Transmission frequencies from animals that have naturally-acquired wild type PT32 *E. coli* O157 strains to in-contact naïve animals.
1.2.3	BIOSS, UOE, MRI	01/09/2015	Statistical analyses of transmission data.
1.2.4	MRI, UOE	01/08/2015	To provide numerical data for modelling under Objective 3.
1.3.1	MRI, UOE	01/03/2016	Excretion and transmission data for a PT21/28 (super-shedding Lineage III) *E. coli* O157 strain no longer containing the Stx2 prophage. Main focus of Stx2a+/- (added 2016)
1.3.2	MRI, UOE	01/05/2016	Measurement of innate and adaptive responses in animals colonised with the Stx2 prophage-excised strain and comparison of these with responses to the Stx2+ parental strain.
1.3.3	MRI, UOE	01/08/2016	Definition of the role of a Stx2 prophage on the excretion dynamics and transmission frequencies of a PT21/28 (super-shedding lineage III) strain. Main focus on Stx2a+/- (added 2016)
2.1.1		30/06/2016	Herd-level prevalence and pat-level counts of *E. coli* O157. Analysis of field studies will provide estimates of herd-level prevalence of *E. coli* O157

	SAC, ADAS, UOE, BIOSS		on farms in Scotland, England & Wales. These estimates will be compared to the estimates generated for the SEERAD and IPRAVE study using published methodology. *E. coli* O157 will be quantified from individual faecal pats. Interim reports for the field studies will be provided on: 31/03/2014, 31/12/2014, 30/09/2015 & 31/03/2016.
2.1.2	SAC/SERL	01/09/2016	Typing and archiving of cattle strains
2.1.3	SAC/SERL	01/09/2016	Provision of EHEC O157 DNA for sequencing
2.1.4	SAC, UOE, BioSS	31/03/2017	Epidemiological analysis of data to answer: (i) Do we see the same association with shedding level/PT/Stx type that was observed in the IPRAVE survey? (ii) How does shedding level relate to strain characteristics determined in 2.1.2?; (iii) what risk factors (food or management related from the survey) are associated with higher shedding?
2.2.1	SAC, PHE, SERL, UoE	31/03/2014	Genome sequences for archived Scottish bovine and human *E. coli* O157 strains from 2002/2003.
2.2.2	PHE, UOE	31/03/2014	Phylogenomic associations between Scottish bovine and human strains from 2002/2003.
2.2.3	UOE, SAC, PHE, SERL	01/09/2016	Genome sequences for bovine and human *E. coli* O157 strains from 2014/15.
2.2.4	PHE, UOE	01/02/2017	Phylogenomic associations of bovine and human strains from 2002/2003 and 2014/15, leading to information about the evolution of bovine and human strains over the last decade in the UK.
2.2.5	PHE, SERL, UoE	01/04/2016	An assessment of the utility and impact of WGS as a routine tool for diagnostic and public health microbiology for *E. coli* O157.
2.2.6	PHE, SERL, UOE	31/03/2017	A standard protocol for sequence analysis of *E. coli* O157 across the UK.

2.2.7	PHE, SERL, UOE	31/03/2017	Enable SERL to transition to sequence-based diagnostics for *E. coli* O157.
2.2.8	PHE, UOE	01/03/2017	Comparative genomic analyses of UK EHEC O157 strains and how they cluster with strains world-wide.
2.2.9	PHE, SERL, UOE	31/03/2017	Confirmation of whether Scotland and/or the UK has an above average incidence of EHEC O157 due to the circulation of specific super-shedding and/or hyper-virulent strains.
3.1.1	MRI, UOE	01/09/2016	Testing of an intervention on excretion and transmission of a representative super-shedding EHEC O157 strain.
3.1.2	MRI, UOE	01/04/2017	Measures of innate and adaptive immune responses of animals in response to an intervention and subsequent challenge with a PT21/28 strain.
3.1.3	BIOSS, UOE	01/05/2017	Statistical analysis of intervention data.
3.2.1	UOG	01/07/2016	The relationship between bacterial shedding density and cattle-to-cattle transmission rates.
3.2.2	UOG	01/02/2017	Predicted impact of interventions on colonisation risk and shedding density.
3.2.3	UOG, UOE, SAC	31/05/2017	A comparison of field intervention strategies and how these impact on the EHEC O157 infection risk to humans.
3.2.4	UOG, UOE, BIOSS, SAC	31/06/2017	Metapopulation models for the transmission of *E. coli* O157 within and between livestock herds that allow the assessment of how interventions with known reduction in prevalence and shedding impact on the risk of human infection
4.1.1	All	30/06/2014	Project meeting with FSA
4.1.2	All	31/12/2014	Project meeting with FSA

4.1.3	All	30/06/2015	Project meeting with FSA
4.1.4	All	31/12/2015	Project meeting with FSA
4.1.5	All	30/06/2016	Project meeting with FSA
4.1.6	All	31/12/2016	Project meeting with FSA
4.3.1	All	30/09/2014	Summary update report on progress against deliverables
4.3.2	All	31/03/2015	Interim annual report 1
4.3.3	All	30/09/2015	Summary update report on progress against deliverables
4.3.4	All	31/03/2016	Interim annual report 2
4.3.5	All	30/09/2016	Summary update report on progress against deliverables
4.3.6	ALL	15/03/2017	Interim report
4.4.1	All	30/06/2017	Submission of plan and outline of final report
4.4.2	All	31/08/2017	Submission of draft final report to FSA
4.4.3	All	30/09/2017	Submission of Final report to FSA

APPENDIX C: British *E. coli* O157 in cattle study manuscript:

[hyperlink to full article]

Epidemiol. Infect. (2017), 145, 3168–3179. © Cambridge University Press 2017
doi:10.1017/S0950268817002151

British Escherichia coli O157 in Cattle Study (BECS): to determine the prevalence of E. coli O157 in herds with cattle destined for the food chain

M. K. HENRY[1], S. C. TONGUE[1]*, J. EVANS[1], C. WEBSTER[1], I. J. MCKENDRICK[2], M. MORGAN[3], A. WILLETT[3], A. REEVES[1], R. W. HUMPHRY[1], D. L. GALLY[4], G. J. GUNN[1] AND M. E. CHASE-TOPPING[5]

[1]Epidemiology Research Unit (Inverness campus), Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK
[2]Biomathematics and Statistics Scotland, James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK
[3]RSK ADAS Ltd., Spring Lodge, 172 Chester Road, Helsby, Cheshire WA6 0AR, UK
[4]Immunity Division, The Roslin Institute, R(D)SVS, University of Edinburgh, Centre for Infectious Diseases, Easter Bush EH25 9RG, UK
[5]Centre for Immunity, Infection and Evolution, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK

Received 7 April 2017; Final revision 2 August 2017; Accepted 30 August 2017;
first published online 19 September 2017

SUMMARY

Escherichia coli O157 are zoonotic bacteria for which cattle are an important reservoir. Prevalence estimates for E. coli O157 in British cattle for human consumption are over 10 years old. A new baseline is needed to inform current human health risk. The British E. coli O157 in Cattle Study (BECS) ran between September 2014 and November 2015 on 270 farms across Scotland and England & Wales. This is the first study to be conducted contemporaneously across Great Britain, thus enabling comparison between Scotland and England & Wales. Herd-level prevalence estimates for E. coli O157 did not differ significantly for Scotland (0·236, 95% CI 0·166–0·325) and England & Wales (0·213, 95% CI 0·156–0·283) (P = 0·65). The majority of isolates were verocytotoxin positive. A higher proportion of samples from Scotland were in the super-shedder category, though there was no difference between the surveys in the likelihood of a positive farm having at least one super-shedder sample. E. coli O157 continues to be common in British beef cattle, reaffirming public health policy that contact with cattle and their environments is a potential infection source.

Key words: Bovine, epidemiology, Escherichia coli (E. coli) O157, estimating disease prevalence.

INTRODUCTION

Human infection with Escherichia coli (E. coli) O157 is a global concern, as infection can lead to kidney failure, neurological complications and haemolytic uraemic syndrome (HUS). HUS can be fatal, particularly in young, elderly or immunocompromised patients [1]. Worldwide, the incidence of HUS due to E. coli O157 infection has been reported at approximately 10% [2], with a 3–5% case-fatality rate [3], while the majority of those who survive suffer some degree of chronic renal function impairment [3]. Cattle and their environments are a reservoir of E. coli O157 [4–6]. Some strains produce verocytotoxin (verocytotoxigenic E. coli (VTEC) O157) and can be excreted in cattle faeces in high numbers, leading to

* Author for correspondence: S. C. Tongue, Epidemiology Research Unit (Inverness campus), Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh EH9 3JG, UK.
(Email: sue.tongue@sruc.ac.uk)

APPENDIX D: Additional data for Figures 2.4 and 2.6

Table to go with Figure 2.4

SEERAD (1998-2000)

Region / location	PT21/28 O157positive	non PT21/28 O157 positive	O157 negative	total sampled
Central	31	15	139	185
Highland	6	9	77	92
Islands	7	8	67	82
North East	39	16	159	214
South East	19	15	101	135
South West	26	23	195	244

IPRAVE (2002-2004)

Region / location	PT21/28 O157positive	non PT21/28 O157 positive	O157 negative	total sampled
Central	6	8	62	76
Highland	7	8	67	82
Islands	7	1	73	81
North East	11	4	66	81
South East	13	6	62	81

South West	**6**	**14**	**60**	**80**

BECS (2014-2015)

Region / location	PT21/28 O157positive	non PT21/28 O157 positive	O157 negative	total sampled
Central	**0**	**3**	**12**	**15**
Highland	**5**	**0**	**15**	**20**
Islands	**4**	**0**	**12**	**16**
North East	**4**	**1**	**14**	**19**
South East	**2**	**3**	**12**	**17**
South West	**2**	**2**	**19**	**23**

Table to go with figure 2.6

BECS: England & Wales

Region / location	PT21/28	PT8	PT54	PT2	PT4	PT32	PT1	PTOther	O157 positive	O157 negative	total sampled
North east	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**3**	**4**	**7**
North West	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**2**	**3**	**9**	**12**
Yorkshire	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**11**	**12**

East Midlands	0	0	0	0	0	0	0	1	1	9	10
West Midlands	1	3	2	0	2	0	2	1	9	16	25
East of England	0	0	1	0	0	0	0	0	1	4	5
London	0	0	0	0	0	0	0	0	0	0	0
South East	0	0	0	1	0	1	0	1	2	13	15
South West	0	1	5	0	0	1	0	1	7	24	31
Wales	1	2	0	0	2	0	2	1	7	36	43

BECS: Scotland

Region / location	PT21/28	PT8	PT54	PT2	PT4	PT32	PT1	PTOther	O157 positive	O157 negative	total sampled
Central	0	1	0	0	0	0	0	2	3	12	15
Highland	5	0	0	0	0	0	0	0	5	15	20
Islands	4	0	0	0	0	0	0	0	4	12	16
North East	4	1	0	0	0	0	0	0	5	14	19
South East	2	1	0	0	1	0	0	1	5	12	17
South West	2	1	0	0	0	1	0	0	4	19	23

APPENDIX E: WGS implementation at SERL and validation
Manuscipt published OP4: **hyperlink here**

Methods

Study Panel: the study panel comprised 110 isolates of *E. coli* O157 and 40 non-O157 *E. coli* isolated between 16/09/14 and 02/02/17. The majority of strains were from human samples but included four non-O157 *E. coli*, from milk (n=1), cheese (n=1) and venison sausages (n=2). The isolates comprising the study panel had been previously characterised using traditional methods (including phage typing, PCR and MLVA) at SERL and represented the diversity of Scottish Shiga toxin-producing *E. coli* causing human infection. Furthermore, the isolates had been previously sequenced and analysed by PHE using validated laboratory procedures and a bioinformatics pipeline to provide identification and typing data, including serotype (O:H type), Multi Locus sequence type (ST), virulence genes (*eae* and Shiga toxin (Stx) subtype) and a SNP address (for the 109 *E. coli* O157:H7), a seven digit profile to describe the clonal population structure. To validate the implementation of WGS at the SERL, the isolates were re-extracted, re-sequenced and the sequence data analysed at the SERL to identify and address any issues hindering the comparison of data between the two laboratories.

Data analysis using the PHE pipeline: the installation at SERL of the Bioinformatics pipeline developed by PHE for variant detection and nomenclature for microbial typing was an important part of this study. This included a piece of software called SnapperDB. Although the pipeline was provided in kind by PHE, its installation required the purchase of a Linux PC. Installation of the PHE pipeline on a Linux PC commenced in July 2016 and was carried out by a BioInformatician PhD student, funded by this Food Standards Agency/Food Standards Scotland collaborative project (FS101055).

Bioinformatics Workflow: the complete workflow was modelled around the PHE WGS workflow. However, an exact copy was not possible due to technical restrictions. The workflow can be divided into four general categories: quality control, typing, variant calling, and databasing. Quality control serves to assure that the data obtained from the sequencer is of high quality and that no issue occurred during the process. The SERL pipeline utilises the known and peer-reviewed tools FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic[76], to obtain read quality metrics and to remove low quality bases or reads.

Typing involves many different steps. The first one aims to determine the content of the sequenced isolate. This is to confirm that what was sequenced is indeed an *E. coli*, and that the sample was not mixed with any other organism. This is done using the PHE developed tool KmerID (https://github.com/phe-bioinformatics). Once the sample is confirmed as an *E. coli* it is put through three genotyping software, all also developed by PHE, named GeneFinder (Doumith unpublished), MLST-typing (new

version named MOST[61], and *stx* subtyping[77]. Each of these tools use read mapping to reference sequences in order to determine presence or absence of virulence and antimicrobial resistance (AMR) genes, the identity of the alleles present for the seven genes making the MLST scheme, and any Shiga toxin genes (*stx*) present. *stx* subtyping also uses an approach which uses BLAST to search the WGS assembly for the *stx* variants (1a-c and 2a-g). This is to determine with great confidence the *stx* subtype, as two key subtypes have as little as 3 amino acid differences between one another. MLST-typing and GeneFinder also provide a predicted serotype.

If the serotype is found to be O157:H7, the sequence proceeds to variant calling. This is achieved by mapping the reads to the O157:H7 reference sequence: Sakai (GenBank accession BA000007). This is done using PHEnix (https://github.com/phe-bioinformatics), a PHE developed tool which relies on peer-reviewed tools bwa[78], a read mapper / aligner, and GATK[79], a variant caller.

Once the variants have been called this is added to the SnapperDB database. This information is then used to create a variant JSON file. This file is a small portable file which contains all the variant information. The portability allows it to be sent to PHE in a rapid manner, so that it is added to their database and a UK wide SNP address is returned to SERL.

While this pipeline is working successfully, and SNP addresses are obtained with a quick turnaround time (~24 hours), there are still many improvements that can be achieved. First, the sending of JSON files to PHE, and the process they have to go through to generate the SNP address could be automated (this is currently being worked on). Secondly, and most importantly, currently the workflow only aims to obtain SNP addresses for the O157:H7 serotype. This was intended; however, this was due to technical restrictions which no longer apply. Therefore, it would be possible to modify the workflow to allow for the generation of SNP addresses for non-O157 *E. coli*.

Sequencing: Sequencing of the 150 strains was conducted over 11 sequencing runs and was completed on 10/03/17.

Results

Concordance Between Laboratories: As detailed in Table A1, excellent concordance (93-100%) was achieved when comparing WGS results produced by SERL with results produced by PHE, demonstrating that standardisation of WGS data between the two Reference Laboratories will be possible.

Table A1. Concordance between the SERL results and those reported by PHE.

Characteristic	Routine [a]	PHE	SERL	Concordance[b] (%)
Species	150	150	150	100
O:H Type	110	150	150	100
ST	-	150	150	100
eaeA	128	127	127	100
Other virulence genes[c]	-	1	1	100
***stx* subtype**	-	150	149	99
SNP address	-	109[d]	101	93

[a] Traditional testing using API20E for *E. coli* species identification; latex agglutination for O157 antigen detection; and real-time PCR for the detection of *eaeA*.

[b] Concordance between the PHE and the SERL results.

[c] *bfpA, aggR, ipaH_type, aaiC, ItcA, sta 1, and stb.*

[d] SNP addresses were only determined for the 109 isolates of *E. coli* O157:H7.

In the eight discrepant cases, only the last digit of the SNP address differed, which means they fall in the same 5 SNP cluster and further analysis showed they differed by either one (n=5), two (n=1), or three (n=2) SNPs. The coverage for the discrepant cases were 41, 66, 72, 83, 87, 98, 118, and 131 suggesting the differences were not related to low coverage.

APPENDIX F: Modelling approach used in Objective 3

1. Modelling

The purpose of modelling is to capture the behaviour of the host-disease system, in order to make predictions about possible scenarios, including evaluating possible responses to various disease control measures.

1.1. Compartmental models

A **compartmental model** categorises different roles in a disease (such as susceptible, infectious, highly infectious animals) into compartments; the level of environmental contamination would be considered its own compartment. The model keeps track of the number in each compartment, the rate at which individuals move between compartments usually through events (such as infection, which converts a susceptible individual into an infectious one), and how the rate at which those events occur changes with time.

1.2. Stochasticity

The low number of individuals in each compartment (there were at most 7 calves involved in each transmission experiment in total), and the need to maintain integer numbers within each compartment requires assuming that each event happens randomly at the appropriate rate. Consequently, every time the model is simulated, a different possible outcome is generated. If the stochastic model is repeated a large number of times, then a distribution of possible outcomes is obtained, which allows the probability of the disease dying out early or of worst case scenarios occurring to be calculated.

1.3. Basic reproduction number

An important metric in disease spread is the basic reproduction number R_0. This metric is defined as *the average number of secondary infectives given introduction of a primary infective into a wholly susceptible population*. If $R_0 > 1$, then the disease has the potential to invade, otherwise the disease is expected to die out quickly (Anderson & May, 1992). R_0 for infectious diseases typically ranges between 1 and 20, with higher R_0 values requiring greater proportions of the population to be vaccinated. The expected proportion of infected individuals at equilibrium (if the disease has no resistant state) is $1 - 1/R_0$, so an R_0 of 3 should equate to $1 - 1/3 = 2/3$ of the population infected.

Diekmann & Heesterbeek[80] provide a method for calculating R_0 algebraically from a specified model.

1.4. Model Fitting

In order for the model to be useful, the set of model parameters controlling the event rates (represented by the vector θ) needs to be adjusted so that its behaviour best captures the behaviour of the underlying system. The model is highly nonlinear, which can make fitting it to the data difficult. Also, given stochasticity, just as the

same set of model parameters can lead to a different set of outcomes, so a different set of parameters could lead to the same outcome. These issues are addressed in Bayesian analysis by obtaining a probability distribution of the model parameters (the so-called “posterior distribution”).

Bayesian inference

Bayes’ Theorem states that:

P(parameters | data) = P(data | parameters) × P(parameters) / P(data)

This means that given some **data**, and any **prior** belief we have about the parameters (which may be uninformative if little is known about the system), we can calculate the **posterior** probability of the model parameters given the data, if we can calculate the probability of the data given the parameters (also known as the **likelihood**), and the probability of the data.

The probability of the data is essentially impossible to calculate. However, this problem can be avoided, since the probability is merely a single normalising constant that does not depend on the choice of parameters, and the posterior can simply be rescaled so that it integrates to 1. This gives an alternative way of representing Bayes’ Theorem:

Posterior ∝ Likelihood × Prior

Finding the posterior distribution θ of the parameters is the main aim of model fitting. In this model, the parameter of prime interest is the vaccine efficacy.

Markov chain Monte Carlo and random walk Metropolis-Hastings

The posterior is a probability distribution over multiple parameters, therefore parameter space may be very large. It is usually inefficient to sample evenly spaced values in order to calculate the posterior, since the posterior is likely densest in a very small subset of the total parameter space (picture a small ball inside a very large room, and measuring every part of the room in order to find the ball). **Markov chain Monte Carlo** (MCMC) is a class of methods for efficiently sampling the parameter distribution θ.

One popular method of implementing MCMC is the **random walk Metropolis-Hastings algorithm** (rwMH) (Hastings, 1970). This algorithm calculates the posterior at a one point p in the parameter space, then proposes a new point in the parameter space q. This new point is accepted with probability

$$\text{Prob(accept)} = \min\left\{1, \frac{\text{Likelihood(data} \mid q) \times \text{Prior}(q)}{\text{Likelihood(data} \mid p) \times \text{Prior}(p)}\right\}$$

This means that if the likelihood of the new point q is more likely than the previous point p, then it is always accepted, however if it is less likely than the previous point,

then it the ratio of likelihoods is less than 1, and the point is accepted with probability equal to this ratio. Consequently, the algorithm explores the parameter space, and tends to move up to regions that are very likely, but can still move away to explore less likely regions such as the tails of the distribution (and allowing it to escape local maxima).

For this method to work, the probability of moving from point p to q must be identical to the probability of moving from point q to point p, so typically the move is small and normally distributed (taking advantage of the symmetry of the normal distribution). It may be the case that a parameter can only be ≥ 0, and a jump might reach a point that is < 0. In this case, the prior distribution states that the probability of that point is zero, and so the new point q is accepted with probability zero (i.e. it is always rejected).

Each successive point p in the chain is recorded (the same point is recorded again if the proposed point is rejected), and when a sufficiently large number of points have been sampled, the chain is examined to find the distribution of each parameter (essentially forming a histogram of the points, and normalising to 1). As the number of points in the chain increases, the distribution of the chain **converges** towards the posterior distribution (this may mean calculating on the order of 10^6 to 10^8 samples, as necessary).

The first few points in the chain are normally discarded, since the starting point may be far from the high-density region, which would over-represent the tails. This set of discarded points is known as the *burn in*, and is often chosen to be the first 10^4 samples or the first 10% of the chain.

Nearby points in the chain are generally autocorrelated (meaning that the probability of being in one point is not independent of being in a nearby point). In order to correct this problem, the chain can be *thinned*, meaning that only 1 point in every 10^4 (say) is chosen, such that adjacent points in the thinned chain are as non-autocorrelated as possible.

Given a chain, it is possible to calculate the *Effective Sample Size* (ESS), which gives the effective number of independent samples after accounting for autocorrelation. Similar to other statistics, higher ESS values are better, and a low ESS indicates that the MCMC algorithm did not sufficiently explore the parameter space to get a good estimate of the posterior distribution (possibly indicating a poor model fit). As a rule of thumb, an ESS value of at least 300 for each parameter is desirable.

Handling missing data: Recorded data is often missing (people were unable to attend that day, measurements were lost, faulty equipment, contamination etc.). In this case MCMC is still able to proceed despite the missing data. Here, each missing value is regarded as an additional latent (“hidden”) parameter, and allowed to vary, just as with any other parameter. The trade-off for being able to calculate the

posterior at all is an increase in the credible interval for each parameter. Where a latent variable is categorical (e.g. the shedding rate for a calf is missing, so it cannot be classified as S, L, or H), then rather than taking a normally distributed jump, one of the different categories is chosen at random. This approach maximises the use of existing data. It is possible that a proposed value for a latent variable is either impossible or highly unlikely, in which case the proposed change is rejected. However, care must be taken at the beginning of the MCMC algorithm to choose starting values that are possible, to prevent the chain from permanently stalling.

1.5. Metapopulation modelling

Metapopulations are populations composed of sub-populations (e.g. all the herds in a given country). By considering multiple herds together, and connecting via movement between herds, the dynamics of a disease can be understood within the metapopulation, which is usually considerably more complex than within a single population in isolation. We account for a range of herd sizes as larger herds may support a higher prevalence due to reduced chance of stochastic extinction, and therefore play a bigger role in maintaining the disease in the metapopulation than smaller herds.

It is important to consider how the movement occurs between groups. A simple model may just assume that movement occurs randomly, between random pairs of herds, and this may be a reasonable way to begin, however it may fail to capture more complex interactions. For example, some herds may be dedicated to breeding new calves, which are then sold to other herds for grazing, before being sold to market. In that case, there is a net movement out of the former herds, and a net movement into the latter herds. Movement may also occur in batches, perhaps seasonally.

The basic reproduction number R_0 is a metric for quantifying the transmissibility of a disease and its potential to invade a population. R_0 may be more difficult to calculate in a metapopulation than in a single population, however if the dynamics within a single herd can be simulated, and the herds in a metapopulation can be reduced to a few categories, then there are efficient methods for examining the transmissibility of a disease within the metapopulation. R_* is the metric analogous to R_0 for a between-groups transmission, which useful to look at between herd spread. When disease transmission is primarily a consequence of movement of infected animals between groups, R_* can be used to calculate the degree of disease control at the point of movement required to prevent disease spread.

APPENDIX G: BECS survey information provided to the participating farms

What were the aims of the project?

- To estimate the percentage of farms in Scotland, which keep cattle for beef, on which *E. coli* O157 could be found.
- To estimate the percentage of farms in England & Wales, which keep cattle for beef, on which *E. coli* O157 could be found.
- To find out whether the *E. coli* O157 that were detected produced Vero toxin, which can cause illness in humans.
- To collect samples of cattle faeces in order to get an idea of all the different types of *E. coli* bacteria that are currently circulating in cattle.

What did we do?

- Visited 110 farms in Scotland and 160 farms across England & Wales.
- Collected faeces samples from the group of cattle that were closest to finishing on each farm on the day we visited.
- Collected information about farm management.
- Tested all samples for the presence of *E. coli* O157.
- Stored a small amount of each sample in our archives.

What did we find?

- In Scotland, we found *E. coli* O157 **in at least one sample** from 26 out of 110 farms, meaning these farms were classified as positive.
- In Scotland, we found Vero toxin-producing *E. coli* O157 on 25 of the 26 positive farms.
- In England & Wales, we found *E. coli* O157 **in at least one sample** from 34 out of 160 farms, meaning these farms were classified as positive.
- In England & Wales, we found Vero toxin-producing *E. coli* O157 on 29 of the 34 positive farms.

All sampled farms

What will we do next?

Bacteria from the samples are being used to study *E. coli* O157 to work out why some of the bacteria will make people very seriously ill, and others will not.

If you gave permission for the samples we gathered from your farm to be used in further studies they may have been used to explore

- Other types of *E. coli* that produce toxins dangerous to humans
- Whether or not there is a link between cattle having *E. coli* O157 and having liver fluke
- Whether the *E. coli* bacteria circulating in cattle are resistant to some antibiotics

About *E. coli* O157

- *E. coli* bacteria are very common in the environment, and many types live in mammals. Some types of *E. coli* cause no problems, but others can lead to human and animal disease.
- *E. coli* O157 is one particular type of *E. coli*. Some *E. coli* O157 produce Vero toxin, which can cause serious illness or death, particularly in very young or elderly people. This type is known as **VTEC O157**.
- VTEC O157 can be carried by cattle and does not appear to affect them in any way.
- You should assume that all cattle might be carrying VTEC O157.
- Proper attention to simple hygiene rules – such as thoroughly washing hands after contact with cattle and their environment – is one of the most effective ways to prevent infection.
- It is also important to keep farm clothing and equipment away from food preparation areas, to prevent cross-contamination.