# Enhanced molecular-based (MLST/whole genome) surveillance and source attribution of *Campylobacter* infections in the UK

# FS101013

https://doi.org/10.46756/sci.fsa.ksj135

Contents


**1. Summary** **6**

**2. Scientific background** **8**

**2.1** What is *Campylobacter*? 8

**2.2** Campylobacteriosis and human infection 8

**2.3** Molecular typing of *Campylobacter* spp. 9

**2.4** Attribution and route of disease 9

**2.5** Antimicrobial resistance 10

**2.6** Human surveillance of *Campylobacter* 11

**2.7** This study 11

**3. Sampling design and descriptive epidemiology** **11**

**3.1** Populations sampled 12

**3.2** Completeness and non-respondents 14

**3.3** Seasonal pattern 20

**3.4** Symptoms reported 21

**3.5** Duration of illness 22

**3.6** Healthcare presentation 23

**3.7** Antibiotic use 24

**3.8** Exposures- food consumed 26

**3.9** Sources of food 27

**3.10** Restaurant, travel, and outdoor environmental exposures 29

**3.11** Ethnicity of respondents 31

**3.12** Conclusion 32

**4. Summary of human *Campylobacter* infection genomic data** **33**

**4.1** Sequencing and assembly validation 34

**4.2** Clonal complex distribution across the study sites and over time 35

**4.3** Comparison of *C. jejuni* across datasets 40

**4.4** *C. coli analysis* 41

**4.5** Conclusion 42

**5. *Campylobacter* detection, enumeration, and characterisation from foods** **43**

**5.1** Methods 43

**5.2** Results 44
**5.3** Discussion 51

**6. Attribution of human disease to source 55**

**6.1** Systematic review of MLST-based source attribution of *Campylobacter* 55
**6.2** Whole genome sequence attribution 58
**6.3** Reference datasets to support attribution 58
**6.4** Validation of attribution using MLST data 63
**6.5** Adjustment for bias in MLST based attribution 78
**6.6** MLST based attribution results, raw and adjusted for identified bias 79
**6.7** Attribution based on sentinel types 86
**6.8** Discussion and conclusion 89

**7. Data visualisation 91**

**7.1** A *Campylobacter* Storyboard 91
**7.2** A *Campylobacter* Dashboard 93

**8. Antimicrobial resistance 95**

**8.1** Method 95
**8.2** Results 95
**8.3** Discussion 101

**9. Cluster detection 102**

**9.1** Analysis of data from this study to inform PHE protocols for cluster detection 102
**9.2** Cluster detection on study data and the PubMLST database using cgMLST 105
**9.3** Conclusions on observed clustering 110

**10. Joint analyses of genomic attribution and epidemiology data 111**

**11. Discussion, future work, and recommendations 113**

**11.1** Sources of human infection 113
**11.2** Methodology to determine the aetiology of human infection 115
**11.3** Antimicrobial resistance and use 116
**11.4** Clustering and outbreaks 117
**11.5** Genome wide association and other basic science studies 117
**11.6** Future sentinel surveillance 118

**11.7** Recommendations 118

**12. References** **121**

**13. Appendices** **132**

**13.1** Appendix 1. Study group 132
**13.2** Appendix 2. Comparison of sequencing at PHE and the Sanger Institute 135
**13.3** Appendix 3. Matching of notifications, questionnaires, and isolates for patient episodes of infection. 146
**13.4** Appendix 4. Additional tables and figures from descriptive epidemiology report. 149
**13.5** Appendix 5. Whole genome sequencing attribution – literature reviews 156
**13.6** Appendix 6. Attribution reference population datasets 159
**13.7** Appendix 7. Datasets from this study 161
**13.8** Appendix 8. Population genetic attribution results in tables. 165
**13.9** Appendix 9. Antimicrobial resistance estimated from genome data 172
**13.10** Appendix 10 Glossary and abbreviations 180

## Study investigators

The study investigators and scientific steering group members are listed in Appendix 1.

## 1. Summary

This human campylobacteriosis sentinel surveillance project was based at two sites in Oxfordshire and North East England chosen (i) to be representative of the English population on the Office for National Statistics urban-rural classification and (ii) to provide continuity with genetic surveillance started in Oxfordshire in October 2003. Between October 2015 and September 2018 epidemiological questionnaires and genome sequencing of isolates from human cases was accompanied by sampling and genome sequencing of isolates from possible food animal sources.

The principal aim was to estimate the contributions of the main sources of human infection and to identify any changes over time. An extension to the project focussed on antimicrobial resistance in study isolates and older archived isolates. These older isolates were from earlier years at the Oxfordshire site and the earliest available coherent set of isolates from the national archive at Public Health England (1997/8). The aim of this additional work was to analyse the emergence of the antimicrobial resistance that is now present among human isolates and to describe and compare antimicrobial resistance in recent food animal isolates.

Having identified the presence of bias in population genetic attribution, and that this was not addressed in the published literature, this study developed an approach to adjust for bias in population genetic attribution, and an alternative approach to attribution using sentinel types. Using these approaches the study estimated that approximately 70% of *Campylobacter jejuni* and just under 50% of *C. coli* infection in our sample was linked to the chicken source and that this was relatively stable over time. Ruminants were identified as the second most common source for *C. jejuni* and the most common for *C. coli* where there was also some evidence for pig as a source although less common than ruminant or chicken. These genomic attributions of themselves make no inference on routes of transmission. However, those infected with isolates genetically typical of chicken origin were substantially more likely to have eaten chicken than those infected with ruminant types. Consumption of lamb's liver was very strongly associated with infection by a strain genetically typical of a ruminant source. These findings support consumption of these foods as being important in the transmission of these infections and highlight a potentially important role for lamb's liver consumption as a source of *Campylobacter* infection.

Antimicrobial resistance was predicted from genomic data using a pipeline validated by Public Health England and using BIGSdb software. In *C. jejuni* this showed a nine-fold increase in resistance to fluoroquinolones from 1997 to 2018. Tetracycline resistance was

also common, with higher initial resistance (1997) and less substantial change over time. Resistance to aminoglycosides or macrolides remained low in human cases across all time periods. Among *C. jejuni* food animal isolates, fluoroquinolone resistance was common among isolates from chicken and substantially less common among ruminants, ducks or pigs. Tetracycline resistance was common across chicken, duck and pig but lower among ruminant origin isolates. In *C. coli* resistance to all four antimicrobial classes rose from low levels in 1997. The fluoroquinolone rise appears to have levelled off earlier and among animals, levels are high in duck as well as chicken isolates, although based on small sample sizes, macrolide and aminoglycoside resistance, was substantially higher than for *C. jejuni* among humans and highest among pig origin isolates. Tetracycline resistance is high in isolates from pigs and the very small sample from ducks.

Antibiotic use following diagnosis was relatively high (43.4%) among respondents in the human surveillance study. Moreover, it varied substantially across sites and was highest among non-elderly adults compared to older adults or children suggesting opportunities for improved antimicrobial stewardship.

The study also found evidence for stable lineages over time across human and source animal species as well as some tighter genomic clusters that may represent outbreaks.

The genomic dataset will allow extensive further work beyond the specific goals of the study. This has been made accessible on the web, with access supported by data visualisation tools.

## 2. Scientific background

*Campylobacter* is the commonest cause of bacterial gastroenteritis among human populations in the developed world, with an estimated 250,000 cases in England and Wales annually (1). As a consequence, campylobacteriosis is recognised as a major public health concern. There are 40 species and subspecies of *Campylobacter* recognised to date (2, 3), of which, *C. jejuni* accounts for 90% of identified campylobacteriosis in humans, with *C. coli* and occasionally *C. lari* and *C. upsaliensis* causing the majority of the remaining cases (4).

### 2.1 What is *Campylobacter*?

*Campylobacter* is a motile Gram-negative, curved rod shaped bacterium and lives as a commensal organism in the gastrointestinal tract of wild, farmed and companion animals (5). *Campylobacter* is micro-aerophilic and doesn't typically grow below 30$^{o}$C, however, factors such as the ability to form biofilms and colonise protozoa (6, 7) mean that *Campylobacter* can also be isolated outside of the gut including from the environment and environmental water. The pathogen can survive common stressors such as aerobic conditions, temperature changes and starvation, to which *Campylobacter* is exposed during food processing and storage (8) (9).

### 2.2 Campylobacteriosis and human infection

There are many potential pathways of human *Campylobacter* infection, but contaminated retail meat products and especially chicken meat have been identified as the most common sources of human campylobacteriosis (10). This may begin on farms where colonisation is common and poor biosecurity and hygiene have been associated with colonisation of chicken flocks (11). *Campylobacter* can survive the processing of meat including slaughter, transport and refrigeration due to characteristics noted above, contributing to the risk of human infection and illness (12). Good food hygiene at home to reduce cross contamination and thorough cooking of meat are considered important to reduce risk of infection (Food Standards Agency). The infectious dose may vary according to strain and host immunity, but is relatively low in the order of hundreds or thousands of cells, and the incubation period generally ranges from one to ten days (13, 14).

Symptoms of campylobacteriosis typically include diarrhoea, vomiting, abdominal pain, headaches and fever (15). National Institute for Clinical Excellence advises antibiotic treatment only for severe or persistent (>7 days) illness and vulnerable individuals (National

Institute for Health and Care Excellence (NICE) - checked 18 Dec 2019). Rarely, more serious complications can occur such as Guillain-Barré syndrome (GBS) (16). High numbers of infections and the prominence of severe diarrhea in developing countries (17) mean that *Campylobacter* is viewed as an important public health hazard by the WHO. It also causes a substantial health and economic burden in industrialised countries, costing the UK around £1bn each year (18).

## 2.3 Molecular typing of *Campylobacter* spp.

Molecular typing methods for *Campylobacter* have been used to differentiate between species, and to identify clusters of strains within species. More recently, genetic sequencing has been used to attribute human clinical disease to underlying animal reservoir sources of infection, to identify clusters, to identify genotypic determinants of resistance to antibiotics in the food chain and predict resistance in human cases. Multi-locus sequence typing (MLST) for *C. jejuni* and *C. coli* targets 7 house-keeping genes, each approximately 500bp in length with each unique nucleotide sequence being assigned an allele number from the PubMLST database (19). This 7 number “allelic” profile is assigned a sequence type (ST) which are then clustered into related groups called ‘clonal complexes’ (CC) named after the central genotype ST (20). With the advent of large-scale whole genome sequencing, core genome MLST (cgMLST) identifies loci which are present within most members of the population (95%), balancing the high resolution needed for a discriminating typing scheme with comparability across a large number of strains (21). The cgMLST scheme includes 1,343 genetic loci based on those that were (i) annotated in the NCTC11168 strain (22), (ii) present across 95% of sequenced isolates in a test dataset, and (iii) excluding identified paralogous loci. In the same way as for traditional seven gene MLST each unique nucleotide sequence is assigned an allele number from the PubMLST database generating an allelic profile across these 1,343 loci. Whole genome MLST (wgMLST) (23), removes the second of these criteria and includes genetic loci missing from a large proportion of other strains in the species. With advances in technology, cgMLST and wgMLST datasets are increasingly available on a large scale through projects such as this.

## 2.4 Attribution and route of disease

Methods of molecular typing, alongside other techniques, can be used to attribute human infection to possible animal sources (24, 25). *Campylobacter* can be isolated from ruminants

(26), pigs (27), wild birds (including ducks, gulls and starlings) (28-30) and also from the environment such as water and beach sand (31-35). MLST is most commonly used due to high reproducibility and comparison between studies using the PubMLST online database. Some clonal complexes (CCs) obtained via MLST are associated with certain sources; for example, ST-21 and ST-45 CC are widespread, multi-host bacterial strains, whereas ST-61 CC is strongly associated with ruminants, ST-257 CC with chicken and ST-628 CC with wild bird sources (36). These previous studies have focused on the use of a 7-locus MLST scheme. As large whole genome sequenced reference datasets (sequenced from isolates of known origin across possible sources) become available attribution using cgMLST or wgMLST data may provide more accurate estimations. Accurately quantifying the relative contribution of each host species reservoir to human infection is important in understanding *Campylobacter* epidemiology and monitoring the impact of control.

Consumption of raw or undercooked meat contaminated with *Campylobacter* spp. has been recognised as one of the primary routes of human disease (37). Recent attribution studies estimate that between 44% and 83% of human disease originates from chicken meat (38). Many factors have also been associated with altered risk of *Campylobacter* spp. infection including seasonality (39-41), exposure within the food chain (42), travel habits of patients (43) antibiotic use (44) and exposure to raw meat and poor drinking water quality (45).

## 2.5 Antimicrobial resistance

Antibiotic treatment of campylobacteriosis is only advised for vulnerable individuals and for severe or persistent illness under guidance from (National Institute for Health and Care Excellence (NICE)), as most patients recover without any specific treatment. Macrolides are the long-established drugs of choice to treat campylobacteriosis (46) when clinically appropriate, with fluoroquinolones as an alternative. Similar antibiotics have been used in agriculture, with strong evidence to suggest that collectively these have led to the emergence of resistant *Campylobacter* spp. strains (47). In the USA, the levels of fluoroquinolone (FQ) resistant *Campylobacter* rose from 1.3% in 1992 to 40.5% in 2001 (48) and an increase in prevalence of macrolide-resistant *C. jejuni* and *C. coli* has been reported (47). *C. coli* strains are more likely to be resistant to antimicrobials than *C. jejuni* (49, 50). Fluoroquinolone resistance levels in food samples from Australia, which does not report agricultural use, are much lower (51).

### 2.6 Human surveillance of *Campylobacter*

The scale of human *Campylobacter* infection and range of possible sources motivates the need for surveillance studies and source attribution investigations to monitor infection, quantify the importance of major sources and support disease control. However, the high infection rate makes detailed surveillance of all cases and sources difficult, therefore sentinel surveillance is used to support source attribution and antibiotic resistance monitoring to monitor human disease and inform control strategies.

### 2.7 This study

In this study a sentinel surveillance approach was undertaken. The principal aim was to estimate the contribution of the main sources to human infection and to identify any changes in attribution over time. Epidemiological questionnaires and genome sequencing of isolates from human cases from October 2015 to September 2018 was accompanied by sampling and genome sequencing of isolates from possible food animal sources. Retail meat sampling and isolate sequencing (duck, sheep and cattle as part of this project and chicken in a separate ongoing FSA retail chicken meat survey) within the study time period was complemented by genome sequencing of isolates from prior chicken and pig abattoir studies. Available historical human surveillance data from Oxfordshire was also analysed for comparison. The emergence of antimicrobial resistance in human isolates was estimated from genome data in the main study and through whole genome sequencing of additional archived human origin isolates. These isolates were from earlier years at the Oxfordshire site (2003/4 and 2006/7) and the earliest available human disease isolates from the national archive at Public Health England (1997). The same genomic assay of antimicrobial resistance was applied to food animal isolates and results compared across species.

## 3. Sampling design and descriptive epidemiology

This section describes the study populations for epidemiological and genomic data, and reports the descriptive epidemiological data based on patient questionnaire responses to support interpretation of the later sections.

## 3.1 Populations sampled

The human sentinel surveillance study included cases with isolates submitted to NHS Trust laboratories in the study areas between 1 October 2015 and 30 September 2018. Two geographically separate sites were selected. These were selected to ensure a mix of urban and rural populations that approximately matched the overall England and Wales population by the Office for National Statistics urban and rural classification, one in Oxfordshire and one in the North East of England (Newcastle, North Tyneside and adjacent parts of Northumbria within the study laboratory catchment areas). Additionally, inclusion of the Oxfordshire site allowed continuity with earlier work in this population where genetic surveillance has been ongoing since late 2003. Table 1 shows the populations served in comparison to national data for the Office for National Statistics urban / rural classification by Middle Layer Super Output Areas (MSOA). Isolate collection ran from October 2015 to September 2018.

Table 1. Description of the populations served by study laboratories according to the Office for National Statistics urban-rural classification (MSOAs - Middle Super Layer Output Areas) for the North East and Oxfordshire as well as sites combined and England and Wales.

| **Urban/Rural classification** | **North East MSOAs** | **North East Population (%)** | **Oxfordshire MSOAs** | **Oxfordshire Population (%)** | **Combined %** | **England and Wales %** |
|---|---|---|---|---|---|---|
| Rural town and fringe | 8 | 61,053 (8) | 15 | 109,191 (17) | 12 | 9 |
| Rural town, fringe in a sparse setting | 2 | 19,174 (2) | 0 | 0 (0) | 1 | 0 |
| Rural village and dispersed | 2 | 16,167 (2) | 16 | 111,789 (17) | 9 | 7 |
| Rural village and dispersed in a sparse setting | 6 | 46,149 (6) | 0 | 0 (0) | 3 | 1 |
| Urban city and town in a sparse setting | 2 | 16,517 (2) | 0 | 0 (0) | 1 | 0 |
| Urban city and town | 17 | 134,522 (17) | 55 | 432,818 (66) | 39 | 45 |
| Urban major conurbation | 62 | 503,424 (63) | 0 | 0 (0) | 35 | 34 |
| Urban minor conurbation | 0 | 0 (0) | 0 | 0 (0) | 0 | 3 |
| Total | 99 | 797,006 | 86 | 653,798 | - | - |

The North East site comprised the catchment populations of the Freeman Hospital, and the North Tyneside NHS Trust. Questionnaires were sent to each patient for whom one or more samples cultured positive for *Campylobacter* during the study period, and isolates were referred to the Public Health England reference laboratory for whole genome sequencing. In Oxfordshire local authorities routinely sent questionnaires to all residents with diagnosed campylobacteriosis. This activity was replaced by the study questionnaire in these local authority areas for the duration of the study. Questionnaire responses were shared with the local authorities to support their routine work. Faecal samples testing positive in the Oxfordshire University Hospitals NHS Trust on PCR testing were cultured for *Campylobacter* and isolates referred to the Public Health England reference laboratory for whole genome sequencing. Although the laboratory catchment largely maps to the county this accommodation between the study and usual practices in Oxfordshire meant that some patients with questionnaire data did not have an isolate in the study (if their sample was sent to a laboratory outside the study catchment) and some positive samples were not included in the questionnaire study (if the patient residence was outside Oxfordshire). Details of matching questionnaires and laboratory isolate results are given in Appendix 3.

## 3.2 Completeness and non-respondents

Patient questionnaires were sent to 6,119 cases, 3,355 male (54.8%) and 2,764 female (45.2%) (

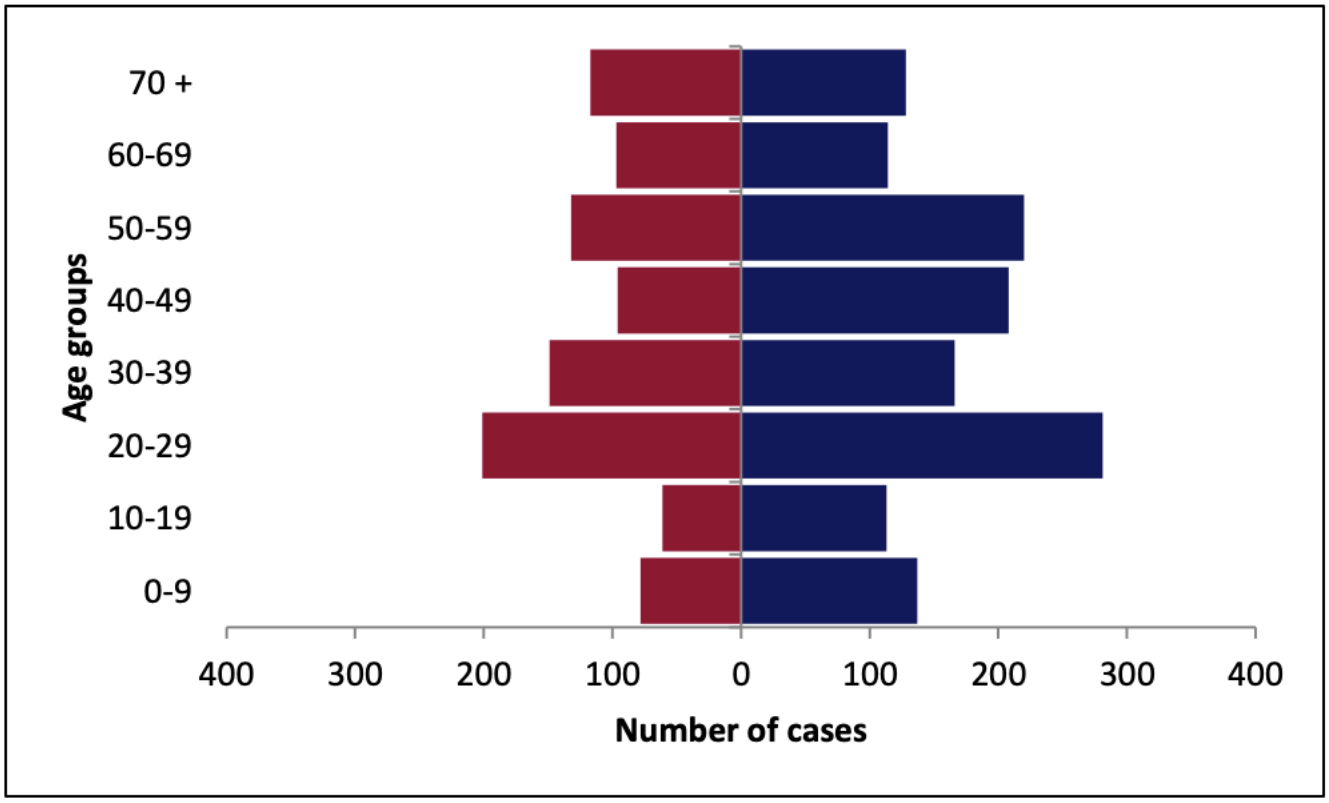
Table 2) in the study areas between 1st October 2015 and 30th September 2018. Across the study 90.5% of cases were sent a questionnaires and reminders within 1 month of isolation excluding those who had replied before the reminder was sent against a target of 90%. In total 3,821 (62.4%) returned a completed questionnaire. A higher proportion of females (n=1,833, 66.3%) responded than males (n=1,988, 59.3%). Response rates in the North East 67.7%) were higher than in Oxfordshire (56.5%) (Chi square $p < 0.001$) (

Table 2).

The median age of the 6,119 cases was 50, cases in the North East had a slightly older median age of 51 years compared to 48 for Oxfordshire. Response rates were higher among older patients and the median age of respondents was 55 in the North East and 54 in Oxfordshire. Non-respondents were most likely to be aged 20-29 which accounted for 23.5% (n=319) of all non-respondents (

Table 3, Table 4,Table 5,

c)




). The median age of non-respondents was 38 overall, 40 in Oxfordshire and 36 in the North East.

Table 2. Summary of responses to survey letters by study site and gender (n=6,119).

| Response status | Gender | North East (n) | North East (%) | Oxfordshire (n) | Oxfordshire (%) | Total (n) | Total (%) |
|---|---|---|---|---|---|---|---|
| Respondents | Female | 1099 | 49.8 | 734 | 45.4 | 1833 | 48.0 |
| | Male | 1107 | 50.2 | 881 | 54.6 | 1988 | 52.0 |
| | Total | 2206 | - | 1615 | - | 3821 | - |
| Non-respondents | Female | 431 | 40.9 | 500 | 40.2 | 931 | 40.5 |
| | Male | 623 | 59.1 | 744 | 59.8 | 1367 | 59.5 |
| | Total | 1054 | - | 1244 | - | 2298 | - |
| Total | Female | 1530 | 46.9 | 1234 | 43.2 | 2764 | 45.2 |
| | Male | 1730 | 53.1 | 1625 | 56.8 | 3355 | 54.8 |
| | Total | 3260 | - | 2859 | - | 6119 | - |

Table 3. Percentage of respondents and gender and age (n=3,821).

| Age group | Female | Male | Total |
|---|---|---|---|
| 0-9 | 2.4 | 3.7 | 6.1 |
| 10-19 | 2 | 3 | 5.1 |
| 20-29 | 5.4 | 4 | 9.3 |
| 30-39 | 5.1 | 3.8 | 8.9 |
| 40-49 | 5.7 | 5.7 | 11.4 |
| 50-59 | 8.4 | 10.2 | 18.6 |
| 60-69 | 9.1 | 11 | 20.1 |
| 70 + | 9.9 | 10.7 | 20.6 |
| Total | 48 | 52 | — |

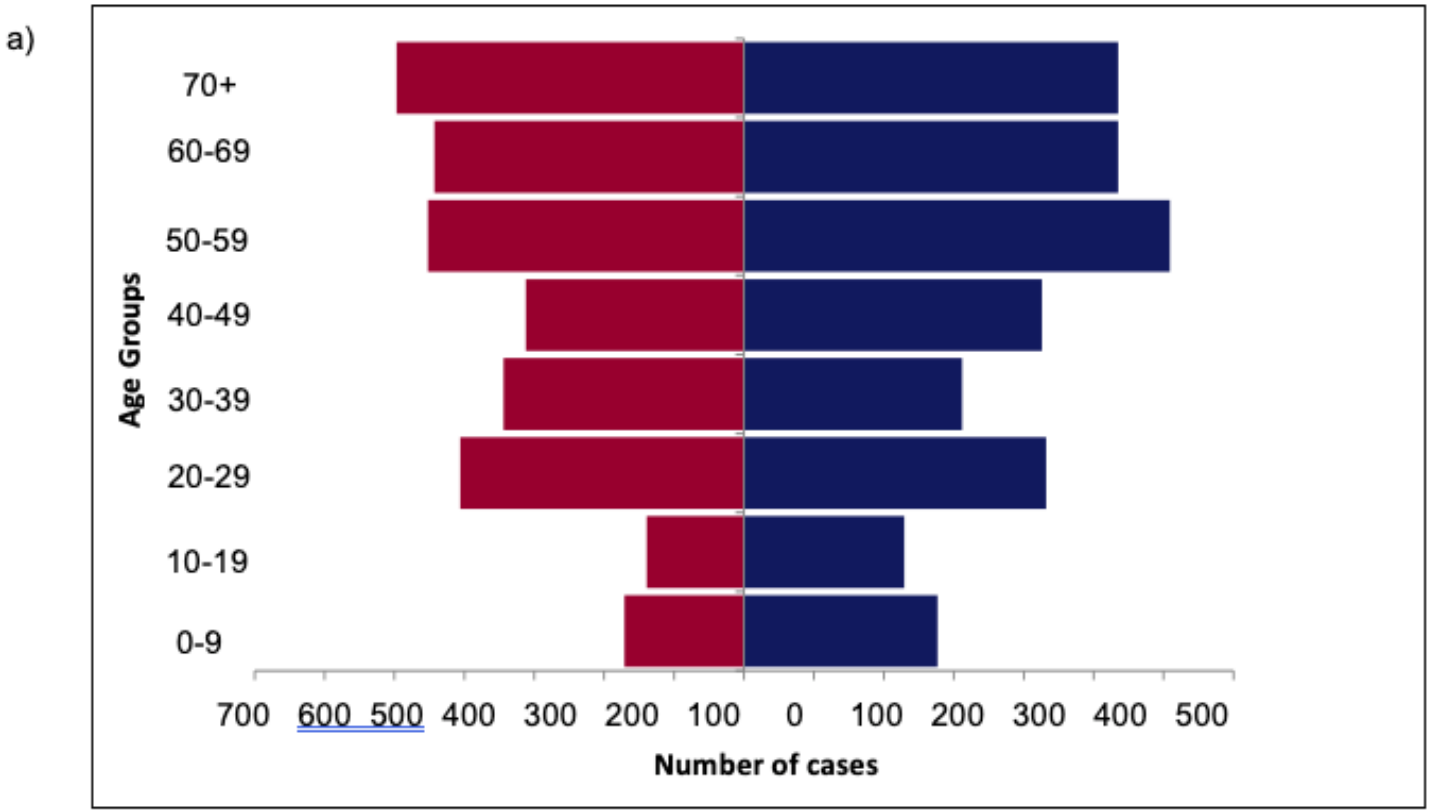
Table 4. Percentage of non-respondents by gender and age (n=2,298).

| Age group | Female | Male | Total |
|---|---|---|---|
| 0-9 | 3.4 | 6 | 9.4 |
| 10-19 | 2.7 | 4.9 | 7.6 |
| 20-29 | 8.7 | 12.2 | 21 |
| 30-39 | 6.5 | 7.2 | 13.7 |
| 40-49 | 4.2 | 9.1 | 13.2 |
| 50-59 | 5.7 | 9.6 | 15.3 |
| 60-69 | 4.2 | 5 | 9.2 |
| 70 + | 5.1 | 5.6 | 10.7 |
| Total | 40.5 | 59.5 | — |

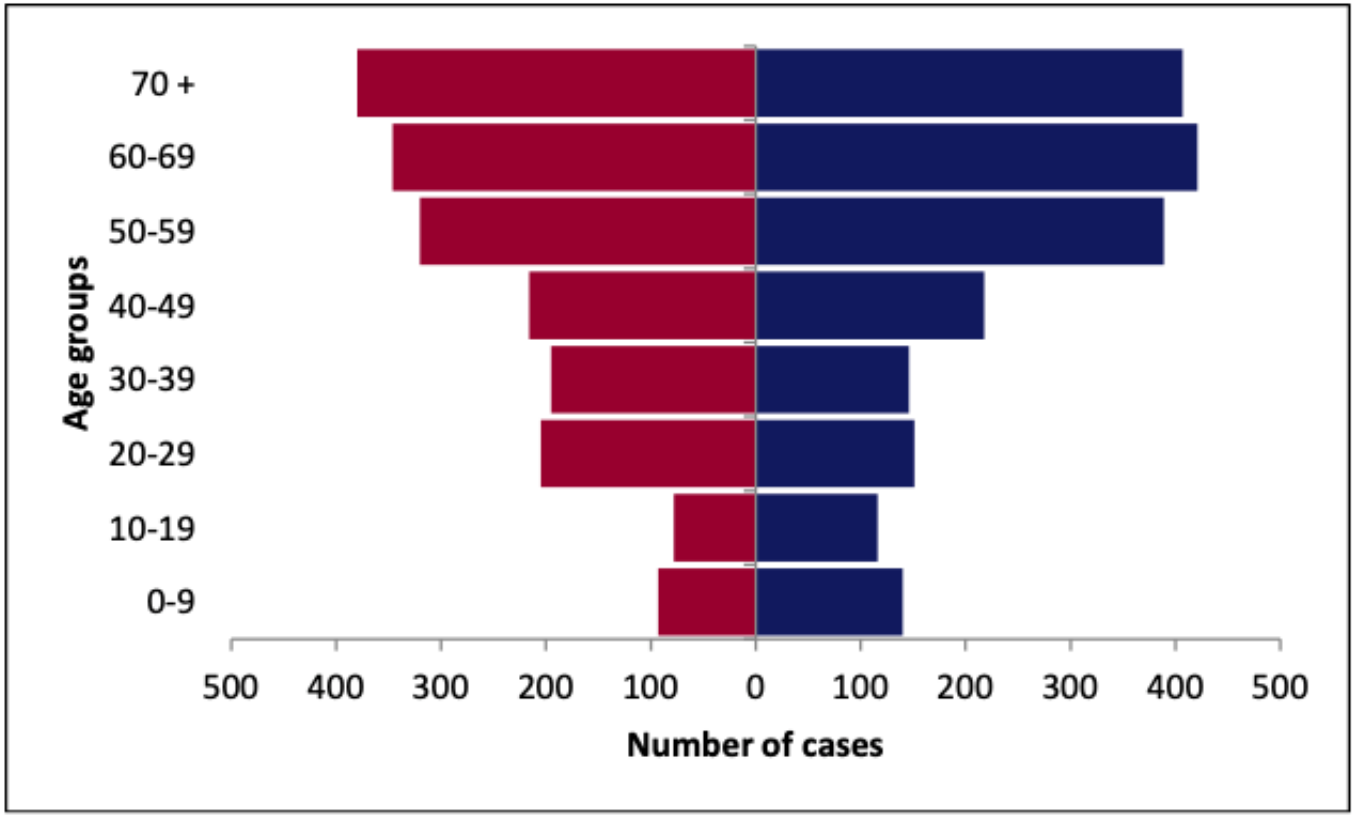
Table 5. Percentage of respondents and non-respondents by gender and age (n=6,119).

| Age group | Female | Male | Total |
|---|---|---|---|
| 0-9 | 2.8 | 4.5 | 7.3 |
| 10-19 | 2.3 | 3.7 | 6 |
| 20-29 | 6.6 | 7.1 | 13.7 |
| 30-39 | 5.6 | 5.1 | 10.7 |
| 40-49 | 5.1 | 7 | 12.1 |
| 50-59 | 7.4 | 10 | 17.3 |
| 60-69 | 7.2 | 8.7 | 16 |
| 70 + | 8.1 | 8.7 | 16.9 |
| Total | 45.2 | 54.8 | — |

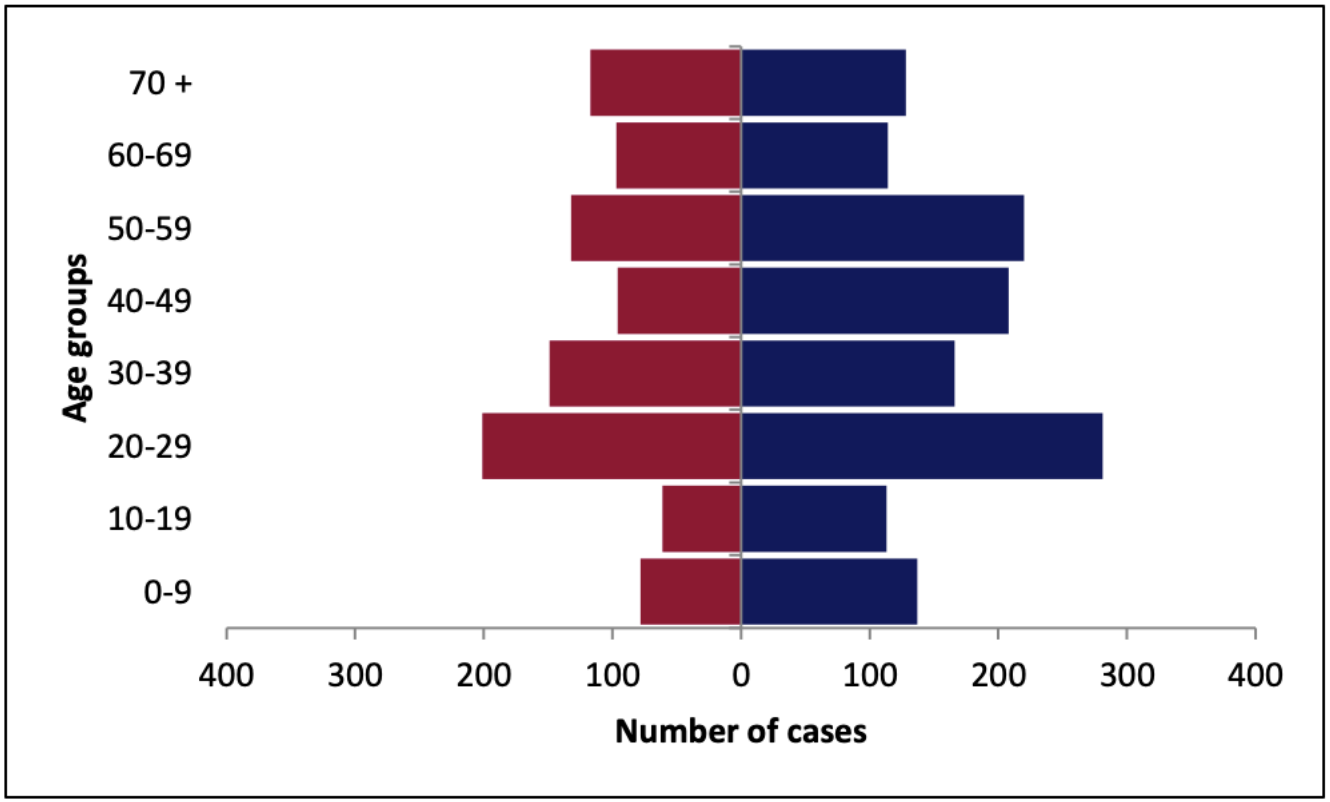
Figure 1. Age-sex distribution of campylobacteriosis cases with female in red and male in blue. a) All cases (n=6,119) b) Respondents (n=3,821) c) Non-respondents (n=2,298).



b)


70 +
60-69
50-59
40-49
30-39
20-29
10-19
0-9
Age groups
500
400
300
200
100
0
100
200
300
400
500
Number of cases

c)


70 +
60-69
50-59
40-49
30-39
20-29
10-19
0-9
Age groups
400
300
200
100
0
100
200
300
400
Number of cases

## 3.3 Seasonal pattern

Overall, the number of cases showed a seasonal peak at both study sites during the summer months followed by a decline over the autumn and winter months (Figure 2) and lowest levels usually in December to February. The summer peak in the North East site was somewhat more distinct and consistent than in Oxfordshire. The total number of cases per study year increased at both sites each year across the study.

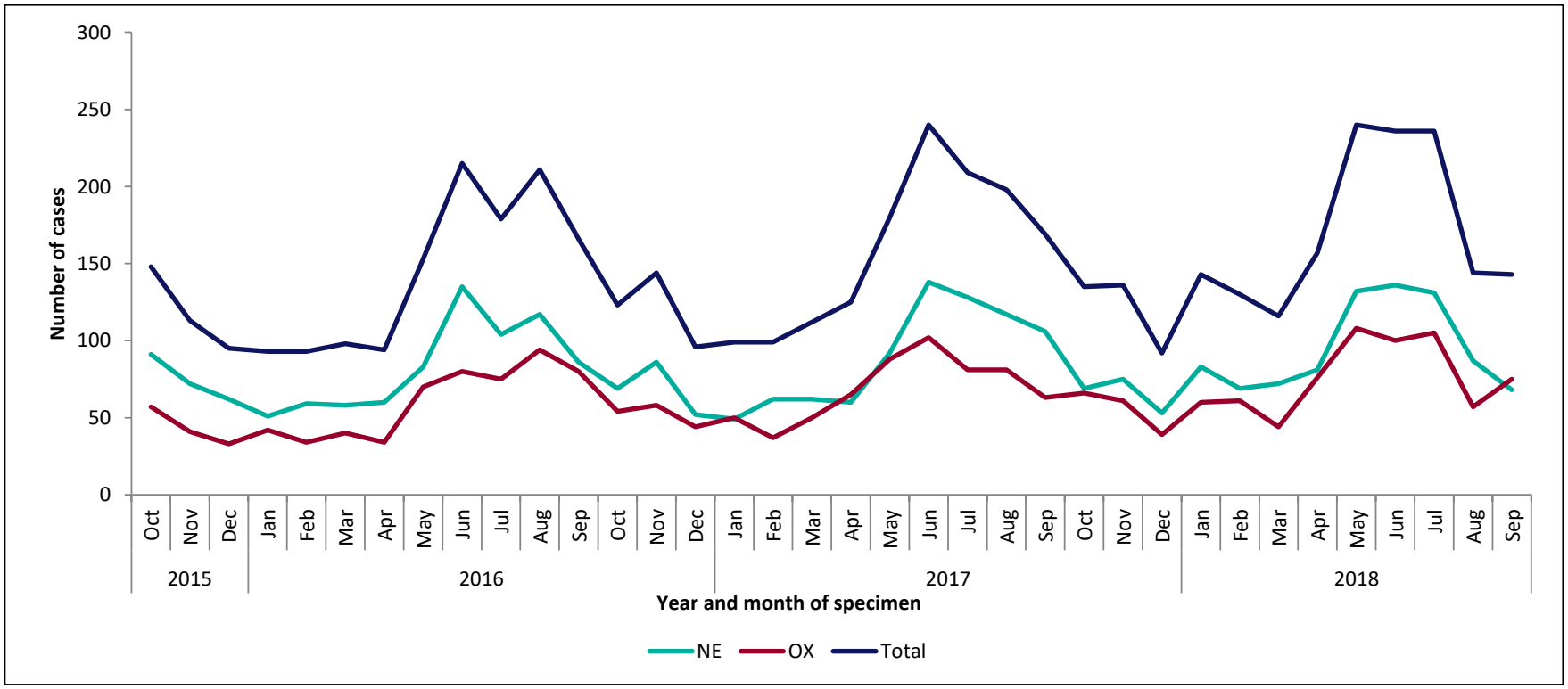



Figure 2. Seasonal pattern of specimen collection date, October 2015 to Sep 2018, overall (n=5,361) and by each study site (North East: n=3,056, Oxfordshire: n=2,305)

## 3.4 Symptoms reported

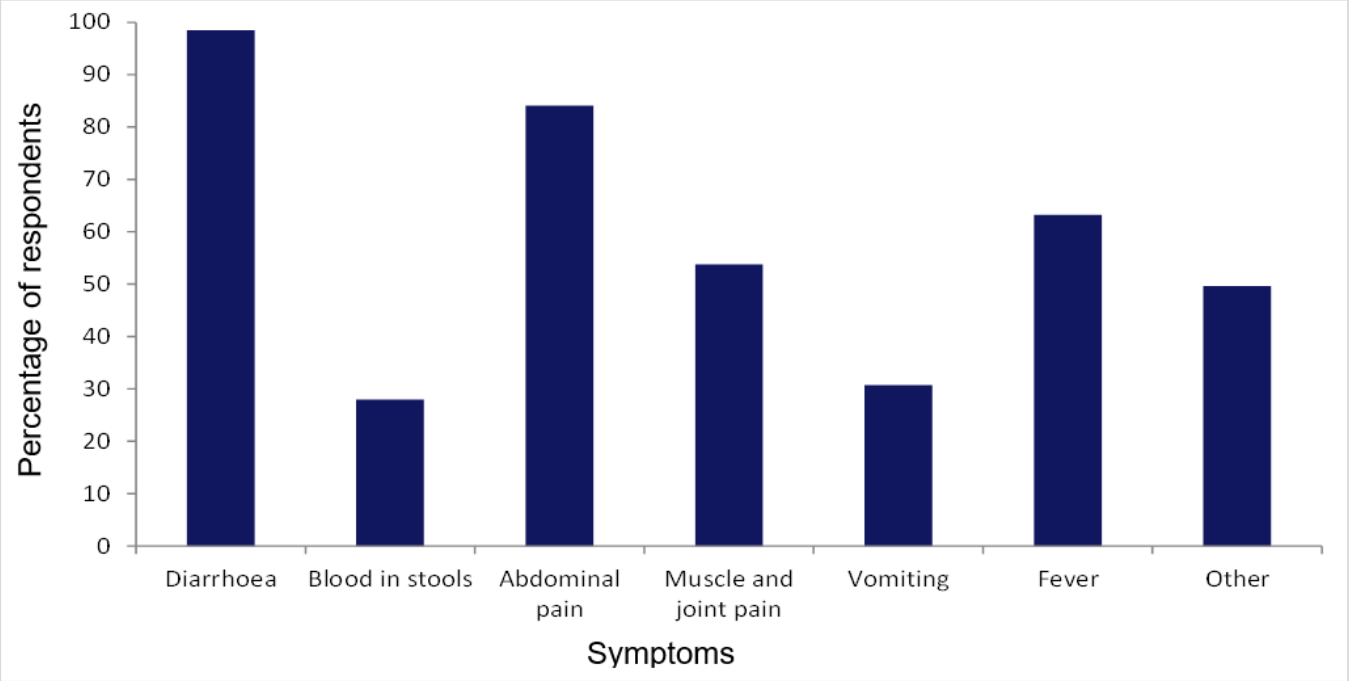



Figure 3. Percentage of respondents reporting a range of symptoms (n=3,816).

Almost all respondents reported diarrhoea (98.5%), 84.1% abdominal pain, and 63.2% fever. Over half the respondents (53.8%) reported muscle and joint pain, 30.7% vomiting, 28.0% blood in their stools, and 49.7% reported other symptoms (Figure 3). Symptom patterns were similar across sites (**Error! Reference source not found.**). Symptoms reported by respondents (n=3,816).

Table 6. The symptoms provided by the respondent exhibited during the infection of campylobacteriosis in the North East, Oxfordshire and across the whole study.

| Symptom | North East (n) | North East | Oxfordshire (n) | Oxfordshire (%) | Total (n) | Total (%) |
|---|---|---|---|---|---|---|
| Diarrhoea | 2175 | 98.7 | 1584 | 98.3 | 3759 | 98.5 |
| Blood in stools | 603 | 27.4 | 465 | 28.8 | 1068 | 28.0 |
| Abdominal pain | 1889 | 85.7 | 1319 | 81.8 | 3208 | 84.1 |
| Muscle and joint | 1225 | 55.6 | 828 | 51.4 | 2053 | 53.8 |
| Vomiting | 689 | 31.3 | 483 | 30.0 | 1172 | 30.7 |
| Fever | 1388 | 63.0 | 1025 | 63.6 | 2413 | 63.2 |
| Other | 787 | 35.7 | 597 | 37.0 | 1895 | 49.7 |

### 3.5 Duration of illness

Of the 3,821 respondents who completed questionnaires 3,100 (81.1%) reported the duration of their illness. A higher proportion of cases at the Oxfordshire site (86.7%) reported the duration than at the North East site (77.1%) but reported patterns of duration were similar. A large proportion reporting durations of 7, 10, 14, and 21 days etc, at both sites presumably reflects rounding of estimated duration of illness to whole weeks and 10 day values in this self-reported retrospective estimation (Figure 4). Reported duration of illness ranged from 1 to 182 days at Oxfordshire and 0 to 180 days in the North East with a median duration of 9 and 10 days respectively and a mean duration of 11 days at each site.

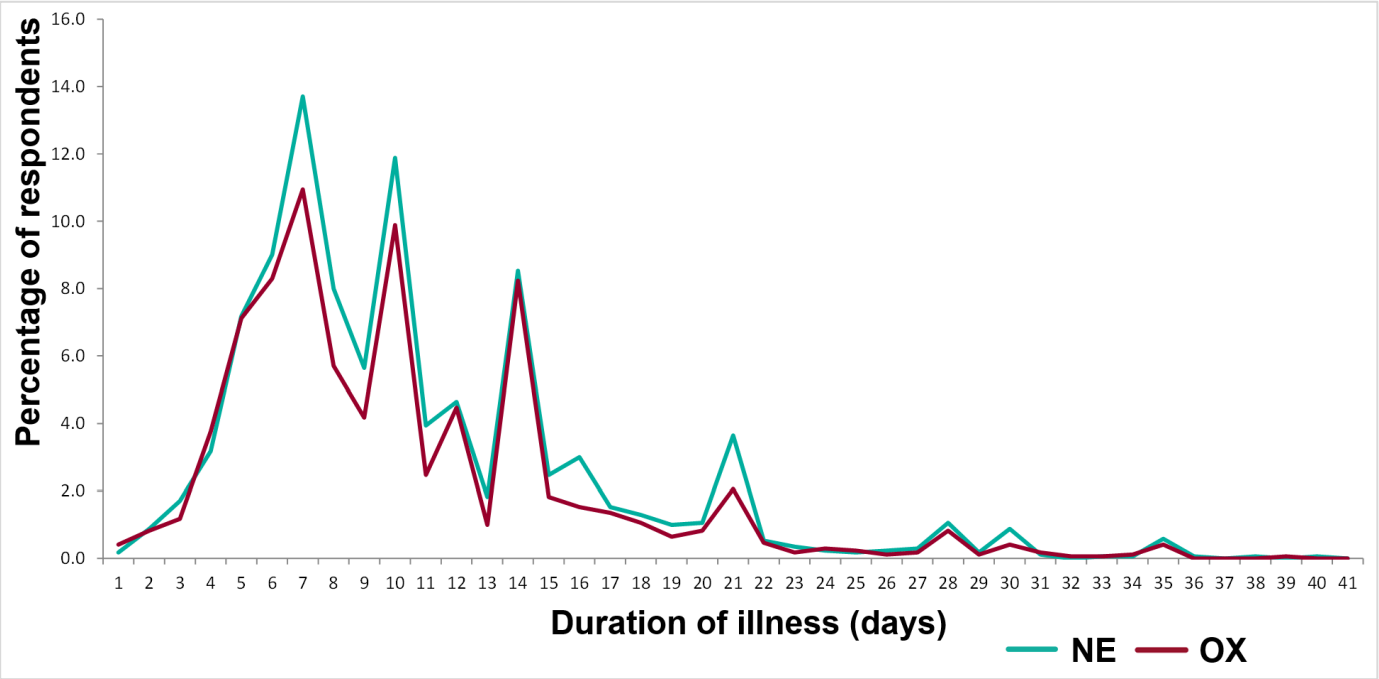



Figure 4. Duration of illness at each study site (NE – North East, OX – Oxfordshire) excluding 28 respondents reporting duration of more than 42 days or longer (n=3,072).

### 3.6 Healthcare presentation

Overall, 90.7% (n=3,465) of respondents reported presenting to their GP for treatment, 23.3% (n=891) presenting at a hospital and 15.6% (n=597) being admitted to hospital (

Table 7). Duration of admission ranged from 1 to 42 days with median 3 days and mean 4.7.

Table 7. Healthcare settings attended by respondents for treatment at each study site and combined (n=3,821).

| Healthcare | North East (n) | North East (%) | Oxfordshire (n) | Oxfordshire (%) | Total (n) | Total (%) |
|---|---|---|---|---|---|---|
| GP | 1987 | 90.1 | 1478 | 91.5 | 3465 | 90.7 |
| Hospital attendance | 544 | 24.7 | 347 | 21.5 | 891 | 23.3 |
| Hospital admission | 330 | 15.0 | 267 | 16.5 | 597 | 15.6 |

### 3.7 Antibiotic use

In the month before symptoms 253 respondents (6.6%) reported antibiotic use, 450 (11.8%) reported taking antibiotics when ill but before their specimen was collected (during), and 1,657 (43.4%) respondents reported use of antibiotics after they provided a stool specimen (Table 8). Reported antibiotic use in the month before symptoms started (6.6%) was similar across the two sites. A higher proportion of respondents reported using antibiotics at the Oxfordshire site compared to the North East site both during (13.1% vs 10.8%, Chi Square p = 0.02) and particularly after specimen collection (53.5% vs 35.9%, Chi Square p <0.001) (Table 8). Missing data was recorded for 8.8% (n=335) of respondents for “use of antibiotics in the month before symptoms started”, 9.9% (n=380) of respondents for “use during specimen collection” and 6.7% (n=202) of respondents for “after specimen collection”.

Table 8. Antibiotic use in the month before symptoms started and during illness before stool specimen collection, and after stool specimen collection among respondents (North East n=2,206, Oxfordshire n=1,615). Missing data ranged from 6% to 11% across these variables.

| North East | Yes (n) | Yes (%) | No (n) | No (%) |
|---|---|---|---|---|
| Before | 144 | 6.5 | 1886 | 85.5 |
| During | 239 | 10.8 | 1765 | 80 |
| After | 893 | 35.9 | 1272 | 57.7 |
| **Oxfordshire** | **Yes (n)** | **Yes (%)** | **No (n)** | **No (%)** |
| Before | 109 | 6.7 | 1347 | 83.4 |
| During | 211 | 13.1 | 1226 | 75.9 |

| | | | | |
|---|---|---|---|---|
| After | 864 | 53.5 | 644 | 39.9 |
| **Total** | **Yes (n)** | **Yes (%)** | **No (n)** | **No (%)** |
| Before | 253 | 6.6 | 3233 | 84.6 |
| During | 450 | 11.8 | 2991 | 78.3 |
| After | 1657 | 43.4 | 1916 | 50.1 |

Table 9. Antibiotic use after specimen collection reported by respondents (n=1,657).

| **Antibiotic** | **North East** | **North East (%)** | **Oxfordshire (n)** | **Oxfordshire (%)** | **Total (n)** | **Total (%)** |
|---|---|---|---|---|---|---|
| Azithromycin | 20 | 2.5 | 15 | 1.7 | 35 | 2.1 |
| Clarithromycin | 80 | 10.1 | 315 | 36.5 | 395 | 23.8 |
| Erythromycin | 240 | 30.3 | 73 | 8.4 | 313 | 18.9 |
| Ciprofloxacin | 64 | 8.1 | 66 | 7.6 | 130 | 7.8 |
| Moxifloxacin | 0 | - | 1 | 0.1 | 1 | 0.1 |
| Amoxicillin/co-amoxiclav | 23 | 2.9 | 15 | 1.7 | 38 | 2.3 |
| Cefazolin/Cefalexin | 5 | 0.6 | 0 | - | 5 | 0.3 |
| Ceftriaxone | 0 | - | 1 | 0.1 | 1 | 0.1 |
| Clindamycin | 1 | 0.1 | 1 | 0.1 | 2 | 0.1 |
| Co-Trimoxazole | 2 | 0.3 | 0 | - | 2 | 0.1 |
| Doxycyline | 58 | 7.3 | 2 | 0.2 | 60 | 3.6 |
| Flucloxacillin | 3 | 0.4 | 0 | 0.0 | 3 | 0.2 |
| Gentamicin | 1 | 0.1 | 1 | 0.1 | 2 | 0.1 |
| Lymecycline | 3 | 0.4 | 0 | - | 3 | 0.2 |
| Meropenem | 1 | 0.1 | 0 | - | 1 | 0.1 |
| Metronidazole | 10 | 1.3 | 10 | 1.2 | 20 | 1.2 |
| Nitrofurantoin | 3 | 0.4 | 0 | - | 3 | 0.2 |
| Penicillin | 2 | 0.3 | 0 | - | 2 | 0.1 |
| Trimethoprim | 2 | 0.3 | 3 | 0.3 | 5 | 0.3 |
| Vancomycin | 0 | 0.0 | 1 | 0.1 | 1 | 0.1 |
| Unknown | 275 | 34.7 | 360 | 41.7 | 635 | 38.3 |
| Total | 793 | - | 864 | - | 1,657 | - |

Of the 253 respondents who reported taking antibiotics in the month before their symptoms started 59.6% (n=151) provided the drug name with amoxicillin-based antibiotics most frequently reported. Of the 450 reporting taking antibiotics while ill but before their stool specimen was collected 150 named the antibiotics taken with amoxicillin-based antibiotics again the most common. Among the larger number of patients who reported taking antibiotics (n=1,657) after their specimen was collected 1,022 (61.7%) named their antibiotic (Table 8). Antibiotic use after specimen collection was dominated by macrolides (44.8%

overall and 69.8% of those with a reported antibiotic name) which were little reported during the earlier periods (13.8% before illness and 7.5% before sample taken). Clarithromycin was the main macrolide used in Oxfordshire and erythromycin in the North East. Ciprofloxacin was the next most frequently reported antibiotic overall (7.8%) and at each study site (Table 8). Overall 51 (1.3%) respondents reported taking more than 1 antibiotic at a given time. Eleven respondents reported taking 2 antibiotics in the month before their specimen was collected, 15 respondents reported taking 2 antibiotics when their stool sample was collected and 32 after they had provided a stool sample. Two respondents reported taking 3 antibiotics after they had provided a stool sample.

National Institute for Clinical Excellence guidance advises antibiotic treatment only for severe or persistent (>7 days) illness and vulnerable individuals. The levels of reported antibiotic use following sampling are high. Those diagnosed may have relatively persistent and severe infection compared to average cases. However these high levels and marked differences across sites may also indicate variable interpretation of, or adherence to, guidelines. Defining antibiotic use defined as starting or switching to an antibiotic recommended for the treatment of campylobacteriosis after stool sample collection, a multivariate model was constructed including factors associated with antibiotic use at $p \leq 0.1$ on univariate analysis, and then using likelihood ratio tests to determine the final model excluding factors not showing statistically significant explanation of antibiotic use within this model. In this multivariable model antibiotic use differed between the sentinel sites (OR 3.10 [2.78-3.88], $p<0.001$). Additionally a higher probability of reported antibiotic use was associated with reported fever (1.40 [1.21-1.74], $p=0.003$), longer periods of illness (2.70 [2.12-3.43], $p<0.001$) and being an adult (18-64) (2.40 [1.63-3.52], $p<0.001$) or elderly (≥65) (1.88 [1.22-2.89], $p=0.004$) compared to those aged under 18. These findings identify GI infections as a potential area for increased attention to antimicrobial stewardship.

### 3.8 Exposures - food consumed

Overall 2,725 (71.3%) of the 3,821 respondents reported eating chicken in the 5 days prior to the onset of their symptoms (Table 10). Of those 37.1% (n=1,416) reported eating chicken at home, 24.8% (n=946) outside their home and 9.5% (n=363) both at and outside of their home. Offal consumption included 3.6% (n=138) reporting eating liver pate or parfait, 3.2% (n=123) lambs liver, 1.8% (n=70) chicken liver, 1.7% (n=65) cow or calves liver, and 0.7% (n=27) pig liver. Consumption of unpasteurised milk was reported by 1.8% (n=67) of

respondents. More detail of the brief food history is available in Appendix 4 along with the study questionnaire. In the absence of control data and given the unreliability of comparing such data with existing food surveys (52) these results are a guide to exposure frequency and are not intended to indicate an association between them and illness.

Table 10. Foods eaten by respondents in the 5 days prior to onset of symptoms (n=3,821) whether at home, or outside, locally, or while travelling.

| Food | North East (n) | North East (%) | Oxfordshire (n) | Oxfordshire (%) | Total (n) | Total (%) |
|---|---|---|---|---|---|---|
| Chicken | 1575 | 71.4 | 1150 | 71.2 | 2725 | 71.3 |
| Duck, turkey or goose | 119 | 5.4 | 116 | 7.2 | 235 | 6.2 |
| Game birds | 10 | 0.5 | 16 | 1.0 | 26 | 0.7 |
| Lamb's liver | 81 | 3.7 | 42 | 2.6 | 123 | 3.2 |
| Cow or calf liver | 25 | 1.1 | 40 | 2.5 | 65 | 1.7 |
| Pig liver | 11 | 0.5 | 16 | 1.0 | 27 | 0.7 |
| Chicken liver | 37 | 1.7 | 33 | 2.0 | 70 | 1.8 |
| Liver pate or parfait | 63 | 2.9 | 75 | 4.6 | 138 | 3.6 |
| Unpasteurised/ raw milk | 36 | 1.6 | 31 | 1.9 | 67 | 1.8 |
| Any cold milk | 583 | 26.4 | 483 | 29.9 | 1066 | 27.9 |

## 3.9 Sources of food

The most frequently reported retailers from which food eaten in the 5 days prior to onset of symptoms was purchased were Sainsbury's (24.8%, n=949) and Tesco (24.5%, n=938), Aldi (18.9%, n=723), Asda (18.3%, n=700) and Morrison's (17.2%, n=657) with fuller details in Table 11. A mean of 28.5% of respondents did not provide a response to the retailer questions with some variation across retailers from 25.0% to 31.1% not marked either yes or no. The overall patterns reported and differences across sites are in the context of differing market share for retailers around the country and consistent with this overall market share (see here). Overall, 32.5% (n=1,243) of respondents reported using loyalty cards with at

least one of the listed retailers, 27.5% at the North East site (n=607) and 39.4% at the Oxfordshire site (n=636). Home-milk delivery was reported by 4.4% (n=169) of respondents and 61.7% (n=2,357) reported eating out in the 5 days prior to the onset of their symptoms.

Table 11. Retailers from which food eaten in the 5 days prior to onset of symptoms was purchased (n=3,821, North East=2,206, Oxfordshire=1,615).

| **Retailer** | **North East (n)** | **North East (%)** | **Oxfordshire (n)** | **Oxfordshire (%)** | **Total (n)** | **Total (%)** |
|---|---|---|---|---|---|---|
| Tesco | 401 | 18.2 | 537 | 33.3 | 938 | 24.5 |
| Sainsbury's | 418 | 18.9 | 531 | 32.9 | 949 | 24.8 |
| Asda | 586 | 26.6 | 114 | 7.1 | 700 | 18.3 |
| Morrison's | 549 | 24.9 | 108 | 6.7 | 657 | 17.2 |
| Waitrose | 103 | 4.7 | 272 | 16.8 | 375 | 9.8 |
| M&S | 289 | 13.1 | 188 | 11.6 | 477 | 12.5 |
| Co-op | 214 | 9.7 | 267 | 16.5 | 481 | 12.6 |
| Aldi | 474 | 21.5 | 249 | 15.4 | 723 | 18.9 |
| Lidl | 218 | 9.9 | 96 | 5.9 | 314 | 8.2 |
| Milk delivered | 101 | 4.6 | 68 | 4.2 | 169 | 4.4 |

## 3.10 Restaurant, travel, and outdoor environmental exposures

Table 12. Summary of reported eating out, travel abroad, and outdoor activities in the 5 days prior symptom onset, with further detail on water exposure during outdoor activity (n=3,821).

| Activity | North East (n) | North East (%) | Oxfordshire (n) | Oxfordshire (%) | Total (n) | Total (%) |
|---|---|---|---|---|---|---|
| Eating out | 1348 | 61.1 | 1009 | 62.5 | 2357 | 61.7 |
| Travel abroad | 387 | 17.5 | 305 | 18.9 | 692 | 18.1 |
| Outdoor activities | 321 | 14.6 | 284 | 17.6 | 605 | 15.8 |
| Freshwater | 66 | 3.0 | 78 | 4.8 | 144 | 3.8 |
| Seawater | 88 | 4.0 | 103 | 6.4 | 191 | 5.0 |
| Other activities | 220 | 10.0 | 191 | 11.8 | 411 | 10.8 |

Table 13 Countries with travel reported by seven or more respondents (n=692).

| Country | Total (n) | Total (%) |
|---|---|---|
| Spain | 174 | 25.1 |
| India | 68 | 9.8 |
| Portugal | 46 | 6.6 |
| France | 39 | 5.6 |
| Turkey | 34 | 4.9 |
| Greece | 25 | 3.6 |
| Morocco | 24 | 3.5 |
| Thailand | 22 | 3.2 |
| Italy | 19 | 2.7 |
| Indonesia | 16 | 2.3 |
| Pakistan | 10 | 1.4 |
| Bulgaria | 10 | 1.4 |
| Poland | 9 | 1.3 |
| Germany | 8 | 1.2 |
| Malta | 8 | 1.2 |
| Cyprus | 8 | 1.2 |
| Malta | 7 | 1.0 |
| Cyprus | 7 | 1.0 |

Countries with 6 respondents (0.8%): Bangladesh, Mexico and USA.

Countries with 5 respondents (0.7%): Cambodia, China, Egypt, Ireland, Malaysia, Netherlands, Nepal, Peru and Vietnam.

Countries with 4 respondents (0.6%): Cuba, Israel, Mauritius, Myanmar, Sri Lanka, Tanzania and UAE.

Countries with 3 respondents (0.4%): Ethiopia and Romania.

Countries with 2 respondents (0.3%): Australia, Austria, Belgium, Canada, Croatia, Czech Republic, Ghana, Hong Kong, Laos, Senegal, Slovakia, Singapore, South Africa, Switzerland, Zambia and Zimbabwe

Countries with 1 respondent (0.2%): Albania, Abu Dhabi, Argentina,  Barbados, Botswana, Bosnia, Cape Verde,  Ivory Coast, Japan, Korea, Latvia, Macedonia, Montenegro, Malawi, New Zealand, Nigeria, Norway, Oman, Slovenia, Uganda, Ukraine and Venezuela. Participation in one or more outdoor activities in the 5 days prior to onset of illness was reported by 605 of 3,821 respondents (15.8%) (Table 12), with 191 (5.0%) respondents reporting seawater activities, 144 (3.8%) freshwater activities and 411 (10.8%) other outdoor activities such as hiking, mountain-biking or camping. Overall 18.1% (n=692) of respondents reported foreign travel to eighty-two different countries, the top 3 destinations were Spain (25.1%, n=174), India (9.8%, n=68) and Portugal (6.6%, n=46) (). Forty-four (6.4%) respondents reported travel to more than one country and 11 (1.6%) didn't report their destination.

### 3.11 Ethnicity of respondents

The study sites selection aimed for national representativeness across the Office for National Statistics urban-rural classification but not across ethnic groups. The self-reported ethnicity in our study group is reported in

Table 14. The 91.7% identifying as “White British” is substantially higher than national data (80.5%). We do not have access to ethnicity for non-respondents. All remaining ethnic groups represented less than 1% of respondents (

Table 14). The questionnaire didn't include any additional "White" (i.e. White Irish, White Gypsy / Traveller) or "Mixed" ethnicity categories. This may have led to some in these groups identifying as "White British". Three per cent of respondents (n=115) didn't identify as any of the listed groups and are reported as "Other" compared to 1% in census data. Given the small numbers in most ethnic groups no analysis by ethnic group is performed and our sentinel results do not allow inference of the epidemiology in individual ethnic groups which would require a larger study and different sampling strategy.

Table 14. Ethnicity of respondents (n=3,821).

| Ethnicity | North East (n) | North East (%) | Total (n) | Total (%) | Total (%) |
|---|---|---|---|---|---|
| Bangladeshi | 14 | 0.6 | 1 | 0.1 | 0.4 |
| Black African | 5 | 0.2 | 4 | 0.2 | 0.2 |
| Black Caribbean | 0 | - | 3 | 0.2 | 0.1 |
| Black Other | 0 | - | 2 | 0.1 | 0.1 |
| Chinese | 7 | 0.3 | 6 | 0.4 | 0.3 |
| Indian | 17 | 0.8 | 12 | 0.7 | 0.8 |
| Pakistani | 16 | 0.7 | 7 | 0.4 | 0.6 |
| White British | 2035 | 92.2 | 1470 | 91.0 | 91.7 |
| Other | 57 | 2.6 | 58 | 3.6 | 3.0 |

## 3.12 Conclusion

This section has described the three-year prospective sentinel surveillance study component of this work that aimed to (i) represent the mix of urban and rural population groups in England within the limits of a two site sentinel study, and (ii) allow continuity with longitudinal sampling in the Oxfordshire site since late 2003. The sample description also identifies the limitations of the study, such as the lack of power to consider differences across ethnic groups.

Findings such as the seasonal pattern observed and the excess of male over female cases are not new and our findings much as expected. Travel is also a well-established risk factor for campylobacteriosis. Among the 18.1% reporting international travel the most commonly visited country (by 25.1% of those reporting travel) was Spain, in line it being the most visited country by UK residents (Statista, visited 17 August 2020). However India, the second most common destination reported by 9.8% of cases describing international travel is a much less frequent destination for visits by UK residents, less than 9% as common as travel to Spain suggesting very a potentially much greater risk from this destination.

The pattern of antimicrobial use described, including the apparently high proportion of patients treated, and the extent to which this varied between the two study sites was striking. Studying antimicrobial use was not a planned objective of the study, with limitations due to

this being self-reported and the associated gaps and potential inaccuracies arising from this, such as in the reported antibiotic names. The findings nonetheless highlight high and variable use of antimicrobials across this population, and the potential importance of self-limiting gastrointestinal infections as an area for antimicrobial stewardship.

## 4. Summary of human *Campylobacter* infection genomic data

This section describes and compares the distribution of *Campylobacter* isolates from human cases across study sites and over time. It also compares the current (2015-2018) data with Oxfordshire data from earlier years (2003/4 and 2006/7) and PHE national archive (1997/8) data sequenced as part of the extension to the main project.

In the current prospective sentinel study 5,325 isolates were received at PHE (3,141 from the North East and 2,184 from Oxfordshire) for sequencing with sequencing completed on 5,024 (94%). Of these, 4,849 (91% of all isolates) were confirmed as *C. jejuni* or *C. coli* with further detail by site in

Table 15 and Table 16. Summary of human study isolates sequenced by PHE.. A high rate of mixed infections identified through sequencing in the early phase of the study was addressed by the introduction of processes to ensure axenic culture at PHE rather than relying on NHS laboratory processes for this.

Table 15. Summary of human study isolates received and sequenced at PHE.

| **All human isolates** | **Oxfordshire** | **North East** | **Total** |
|---|---|---|---|
| Isolates received | 2184 | 3141 | 5325 |
| Isolates sequenced | 2022 (93%) | 3002 (96%) | 5024 (94%) |
| Contaminated (not sequenced) | 140 | 67 | 207 |
| No growth (not sequenced) | 22 | 72 | 94 |

Table 16. Summary of human study isolates sequenced by PHE.

| **Sequenced isolates** | **Oxfordshire** | **North East** | **Total** |
|---|---|---|---|
| Sequence confirmed as *C. jejuni* / *C. coli* | 1955 (90%) | 2894 (92%) | 4849 (91%) |
| *Campylobacter jejuni* | 1773 | 2609 | 4382 |
| *Campylobacter coli* | 182 | 285 | 467 |
| *Campylobacter lari* | 1 | 7 | 8 |
| *Campylobacter upsaliensis* | 0 | 1 | 1 |
| *Campylobacter* spp. mixed | 58 | 96 | 154 |
| Not *Campylobacter* spp. | 8 | 4 | 12 |

## 4.1 Sequencing and assembly validation

Sequencing results from PHE were compared with sequencing results from the Sanger Institute to show consistency as requested by the FSA. This was assessed directly, by comparison of results for isolates sequenced at both PHE and the Sanger Institute before the main project started, and indirectly early in the project by comparing the assembly statistics for the first 309 clinical isolates sequenced in this project, with the values for these metrics from Sanger Institute sequenced Oxfordshire clinical isolates obtained between 2011 and 2015.

In direct comparison 7 locus MLST profiles derived at from Sanger Institute and PHE sequencing were identical for all isolates. Using wgMLST for the 1,553 genetic loci analysable across the full dataset differences between sequencing from the same sample varied between 0 and 12 (0 – 0.08%). A phylogeny drawn using all the results clustered the sets of same patient isolates, sequenced at each centre, tightly together and clearly separated each patient from those of all other patients (Figure 40, Appendix 2). In the indirect comparison, assembly statistics for isolates in the current study were similar to historical data from the Sanger Institute. The number of contigs per sample had a mean of 50 (standard deviation 102) compared to a mean of 51 (standard deviation 97) in past isolates assembled at the Sanger Institute, and assembled genome mean length was 1.70 megabases in each dataset. Full results of these validations are reported in Appendix 2.

## 4.2 Clonal complex distribution across the study sites and over time

In the North East 2,234 *C. jejuni* and 240 *C. coli* isolates were sequenced and in Oxfordshire 1,497 *C. jejuni* and 151 *C. coli* isolates. The clonal complex distribution for *C. jejuni* is similar across the two sites ($x^2$, p=0.06) (Figure 5).

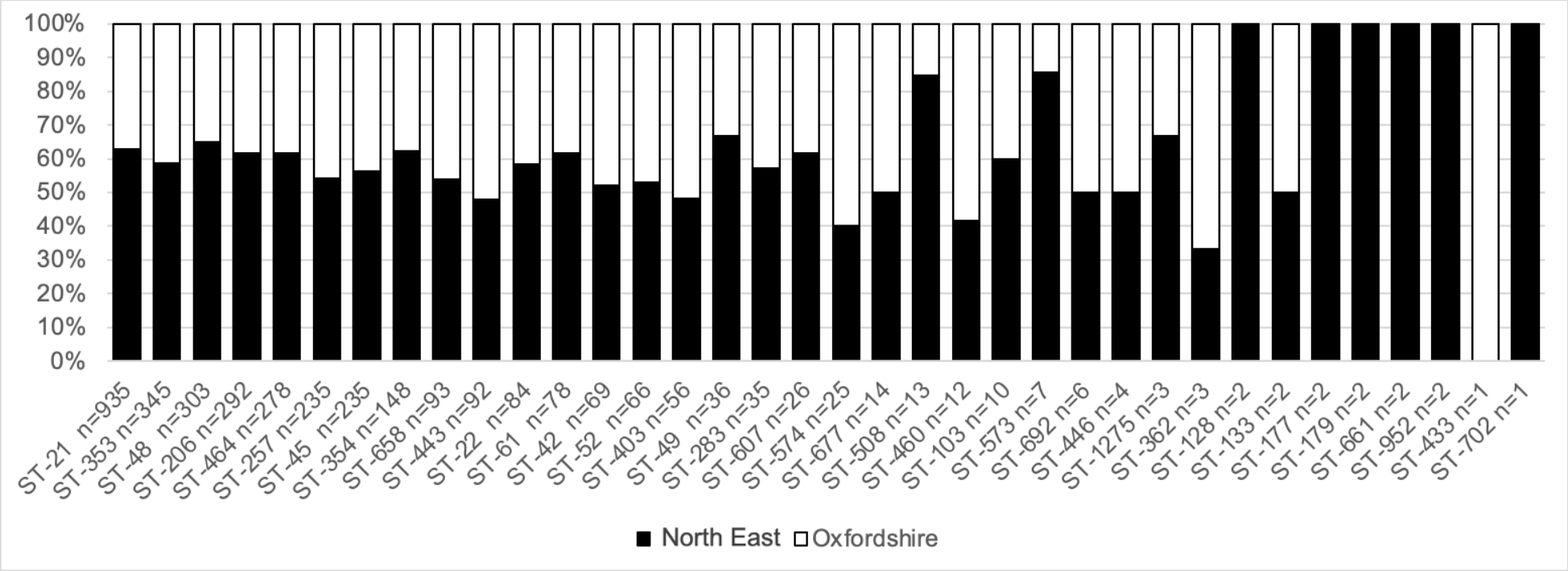
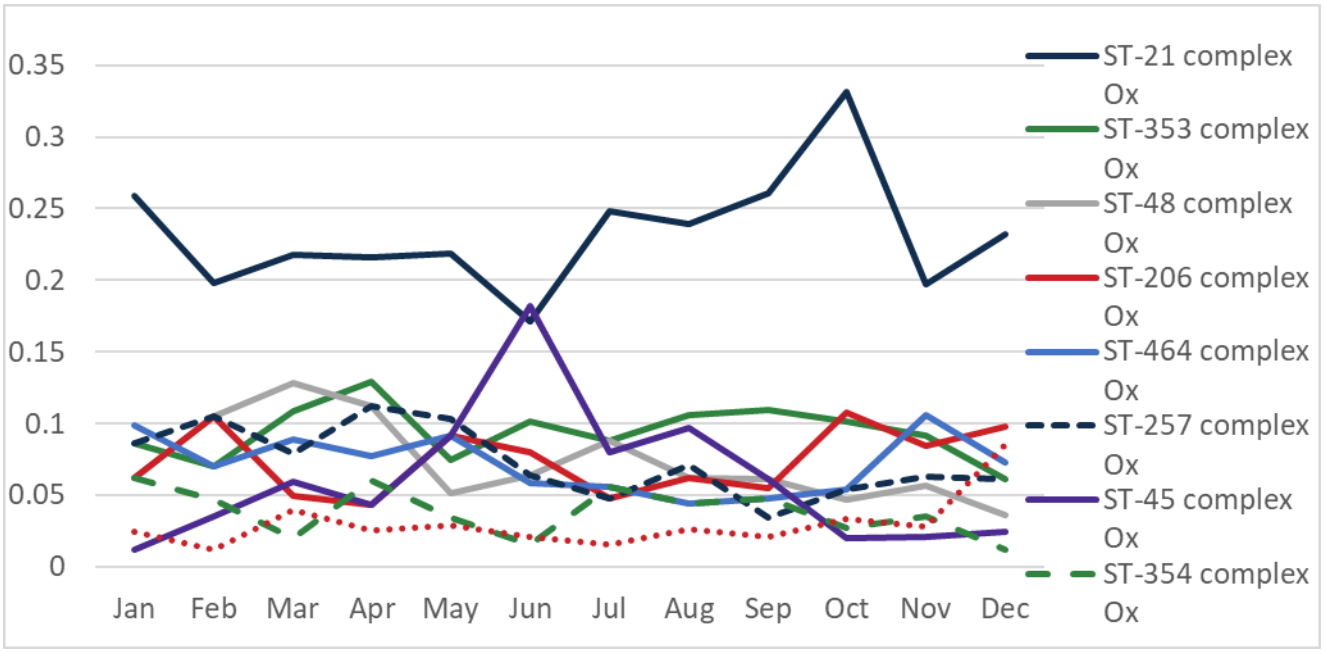
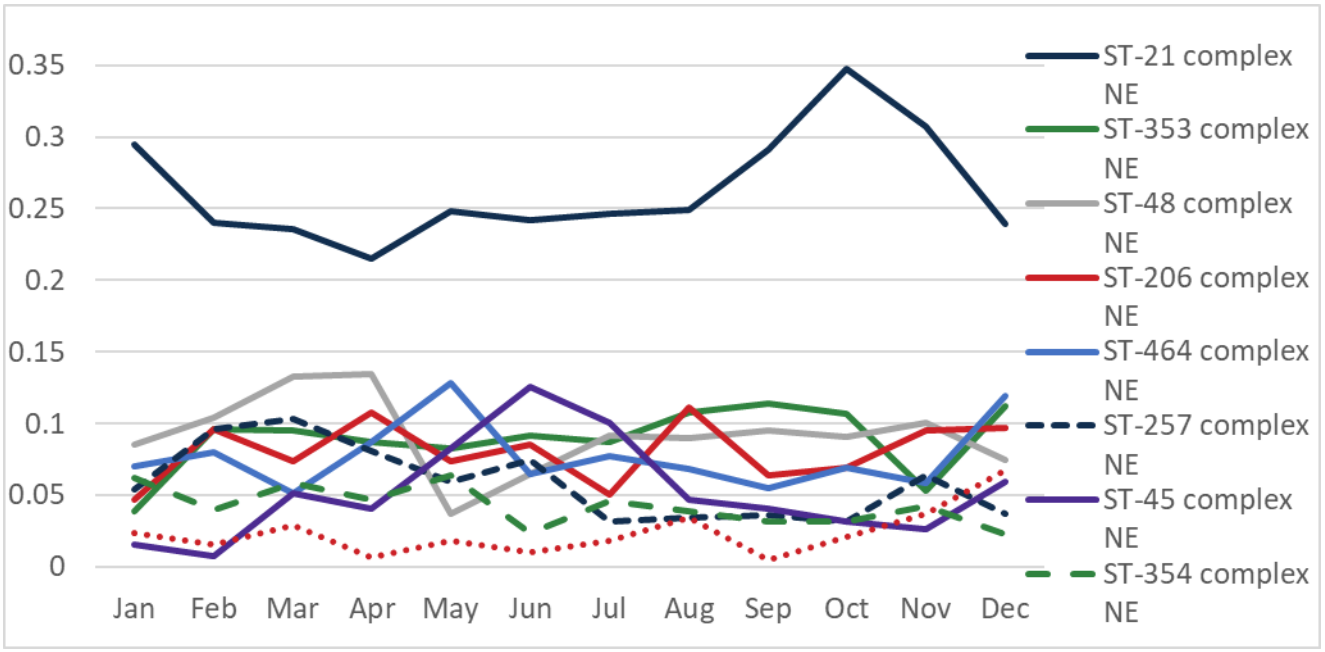



Figure 5: Percentages of *C. jejuni* isolates in each clonal complex originating in the North East and in Oxfordshire presented as a stacked bar graph. Overall, 60% of isolates were from the North East and 40% Oxfordshire. X-axis labels indicate number of isolates per clonal complex, descending in number from left (n = 935 for ST-21 complex) to right.

The distribution across clonal complexes was broadly similar across the three years of the study and across the two sites (
). Both numbers and percentages are presented given that overall numbers dropped in the North East and increased in Oxfordshire across the three years of the study. The relative seasonal contribution for each of the nine most common clonal complexes is shown in

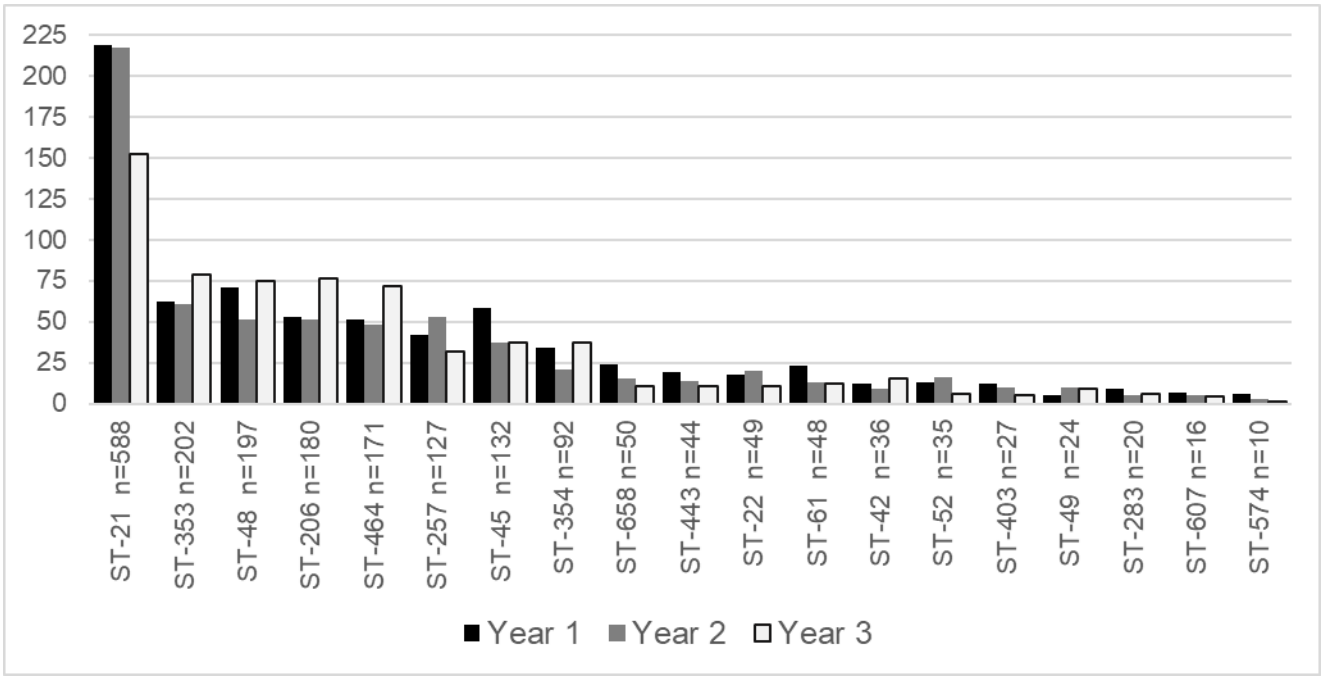


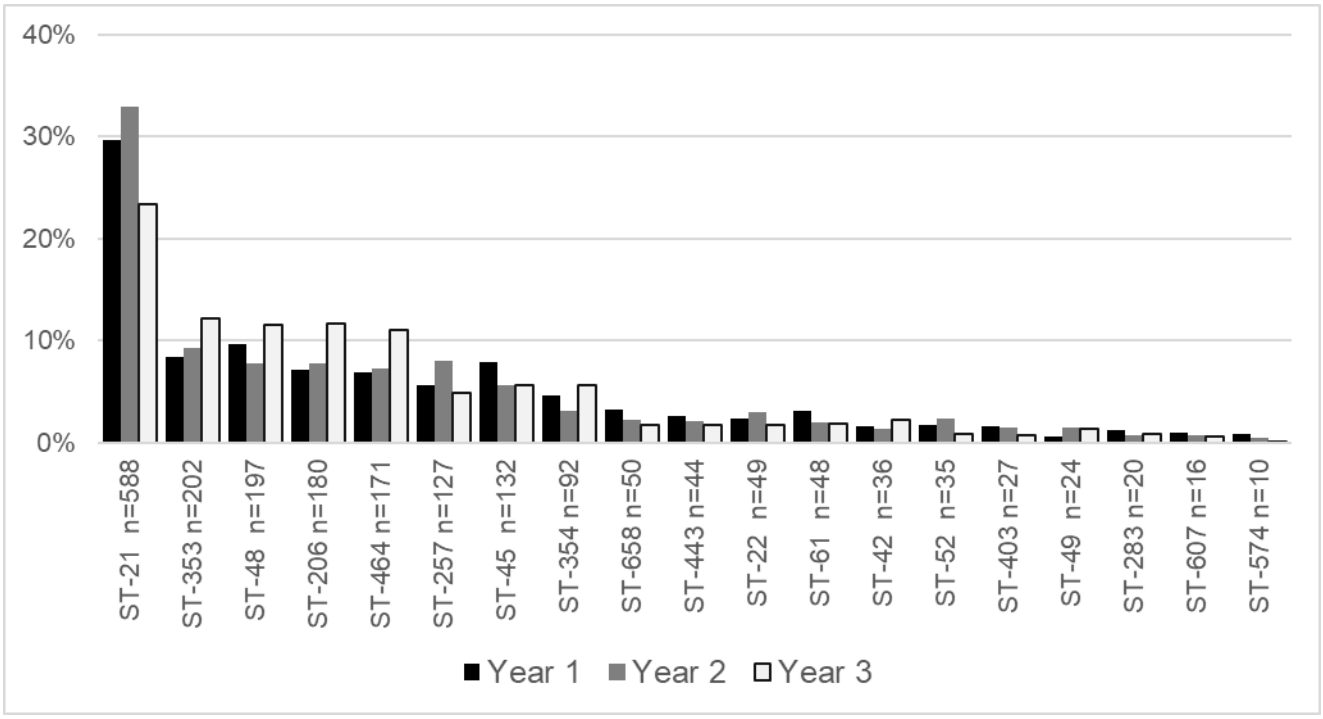
. Patterns are similar across the sites with a summer peak of ST-45 complex the most pronounced feature.

Figure 6: The number and percentage of each clonal complex in the North East (a and b) and Oxfordshire (c and d) by study year (Year 1 -2015/16, Year 2 -2016/2017 and Year 3 -2017/18), for clonal complexes identified 15 times or more. The percentage of isolates from each clonal complex that year divided by the total that year across all of these clonal complexes.

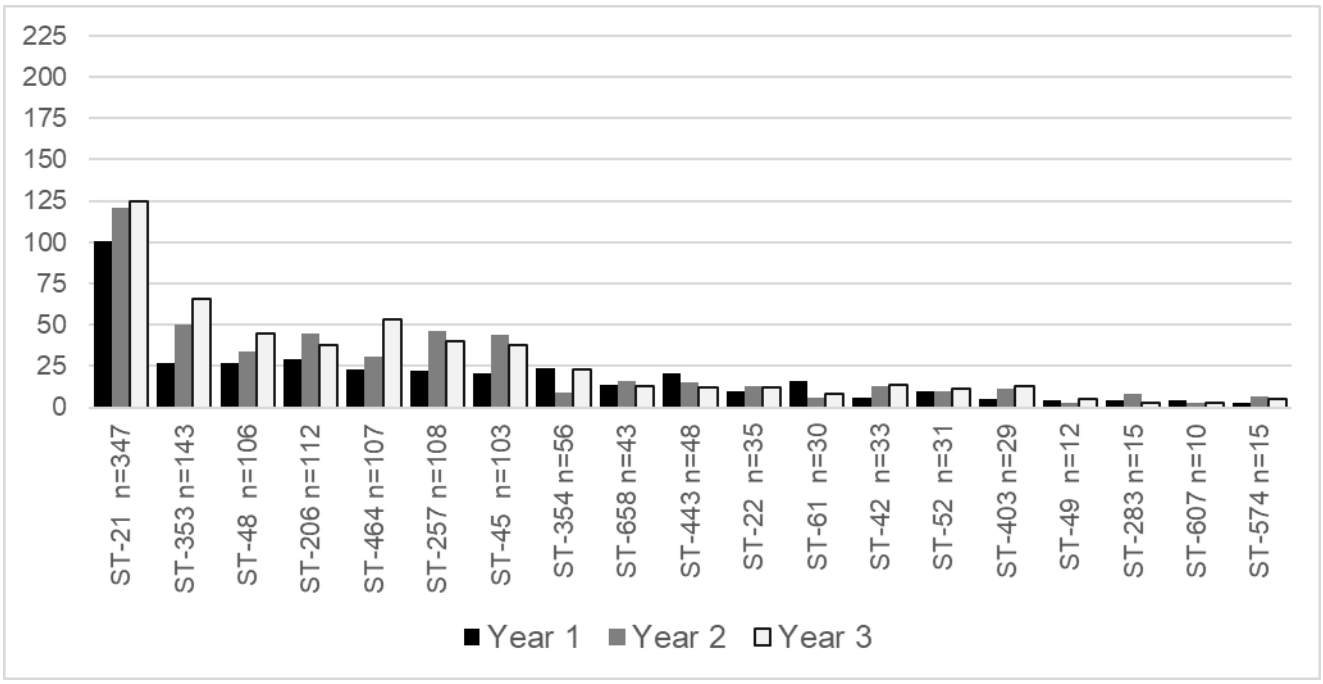
a



b


40%
30%
20%
10%
0%
ST-21 n=588
ST-353 n=202
ST-48 n=197
ST-206 n=180
ST-464 n=171
ST-257 n=127
ST-45 n=132
ST-354 n=92
ST-658 n=50
ST-443 n=44
ST-22 n=49
ST-61 n=48
ST-42 n=36
ST-52 n=35
ST-403 n=27
ST-49 n=24
ST-283 n=20
ST-607 n=16
ST-574 n=10
Year 1
Year 2
Year 3

c


225
200
175
150
125
100
75
50
25
0
ST-21 n=347
ST-353 n=143
ST-48 n=106
ST-206 n=112
ST-464 n=107
ST-257 n=108
ST-45 n=103
ST-354 n=56
ST-658 n=43
ST-443 n=48
ST-22 n=35
ST-61 n=30
ST-42 n=33
ST-52 n=31
ST-403 n=29
ST-49 n=12
ST-283 n=15
ST-607 n=10
ST-574 n=15
Year 1
Year 2
Year 3

d

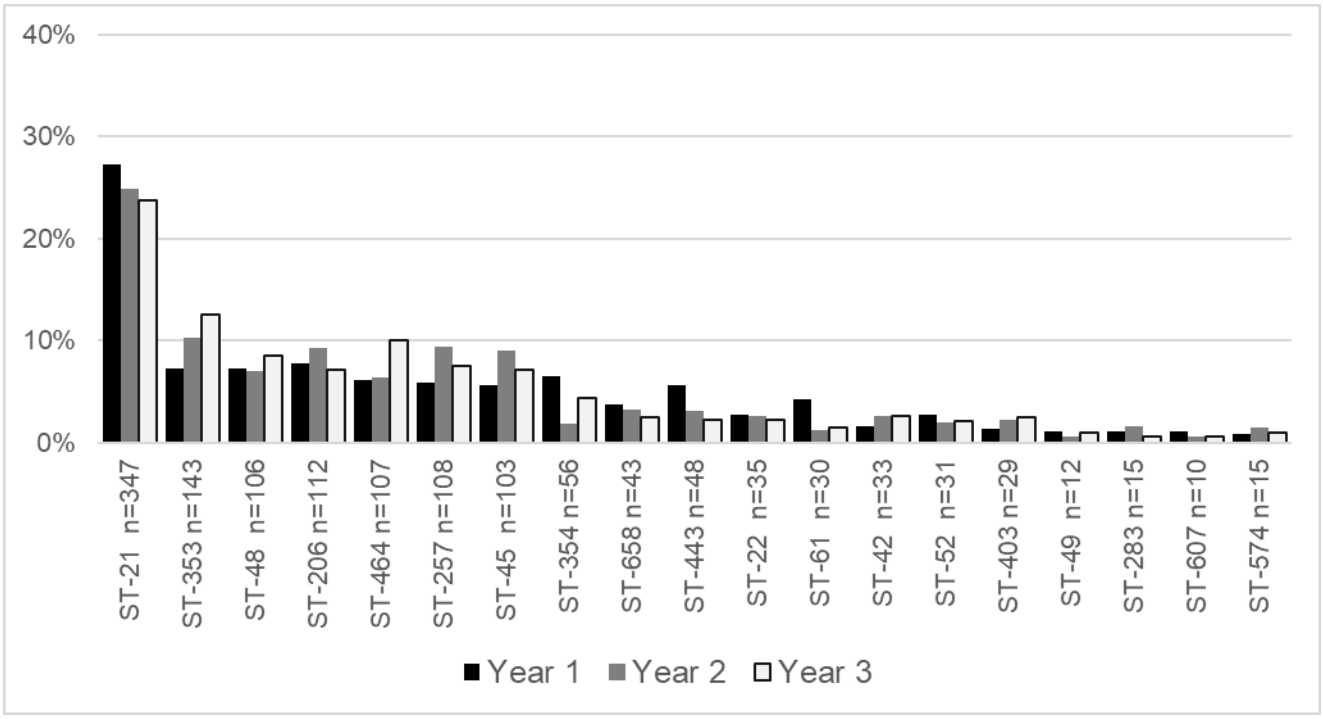
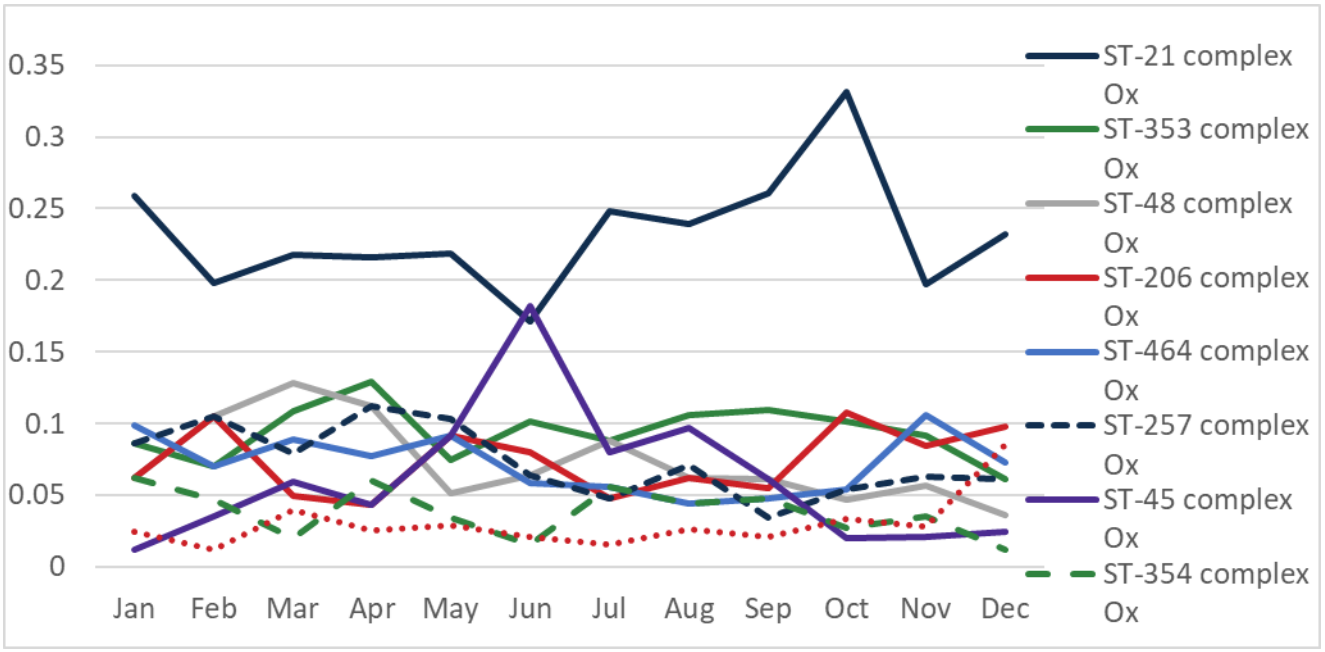
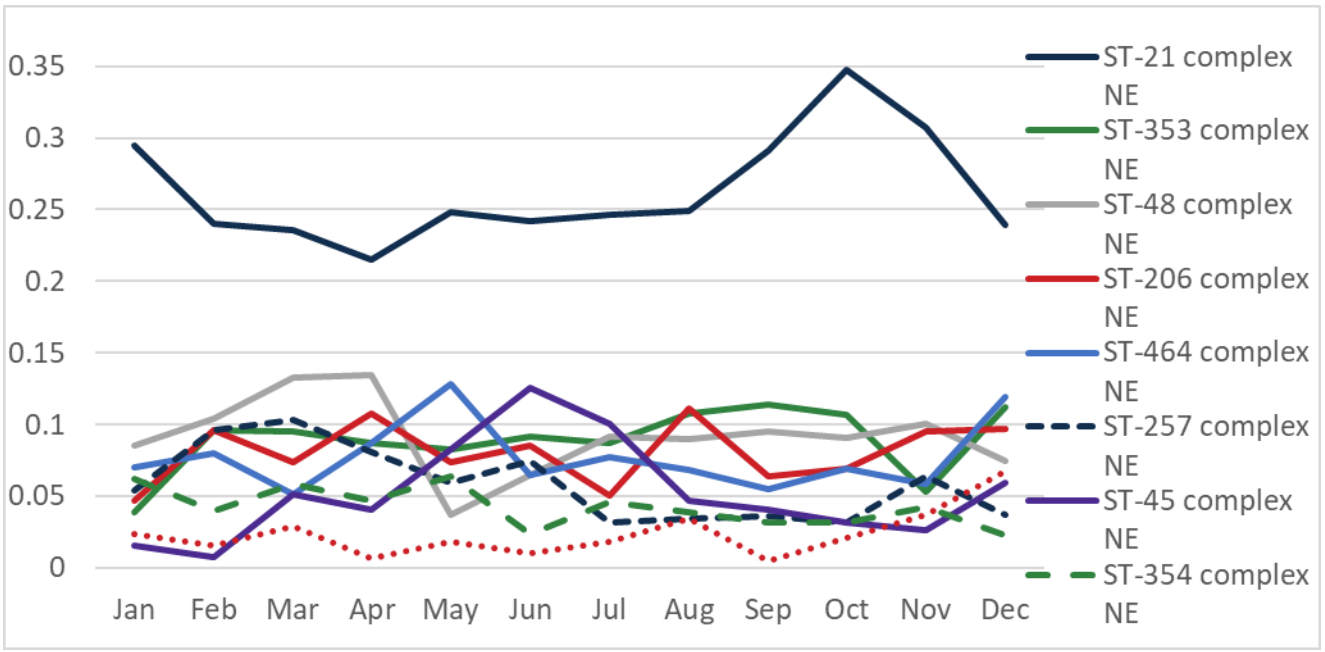

40%
30%
20%
10%
0%
ST-21 n=347
ST-353 n=143
ST-48 n=106
ST-206 n=112
ST-464 n=107
ST-257 n=108
ST-45 n=103
ST-354 n=56
ST-658 n=43
ST-443 n=48
ST-22 n=35
ST-61 n=30
ST-42 n=33
ST-52 n=31
ST-403 n=29
ST-49 n=12
ST-283 n=15
ST-607 n=10
ST-574 n=15
Year 1
Year 2
Year 3

Figure 7: The proportion of all isolates in each of the 9 most common clonal complexes, by month, in Oxfordshire (upper panel) and the North East (lower panel).



## 4.3 Comparison of *C. jejuni* across datasets

The prevalence of each clonal complex was compared across the time periods from 1997/8 to 2015-8 (Figure 8). The data for 1997/8 is from nationally distributed PHE archive data. The two intermediate years run from October to the following September from the Oxfordshire archive (2003/4 and 2006/7) and the current study comprises the data across the three years from October 2015 to September 2018. There was significant variation in clonal complex distribution across these four sampling periods ($X^2$, $p<0.001$). Although ST-21 is the most prevalent clonal complex throughout, its relative prevalence decreases between 1997/98 and the three later time periods. ST-206, ST-353, ST-403 and ST-464 increase in relative prevalence over this time period, with ST-464 apparently emerging between 1998 when and 2003 and increasing in more recent years. ST-257, ST-354, ST-443 and ST-574 complexes all show a pattern of rising after 1997/8, peaking during one of the two time middle periods (2003/04 and 2006/07), and decreasing during the most recent sentinel study period.

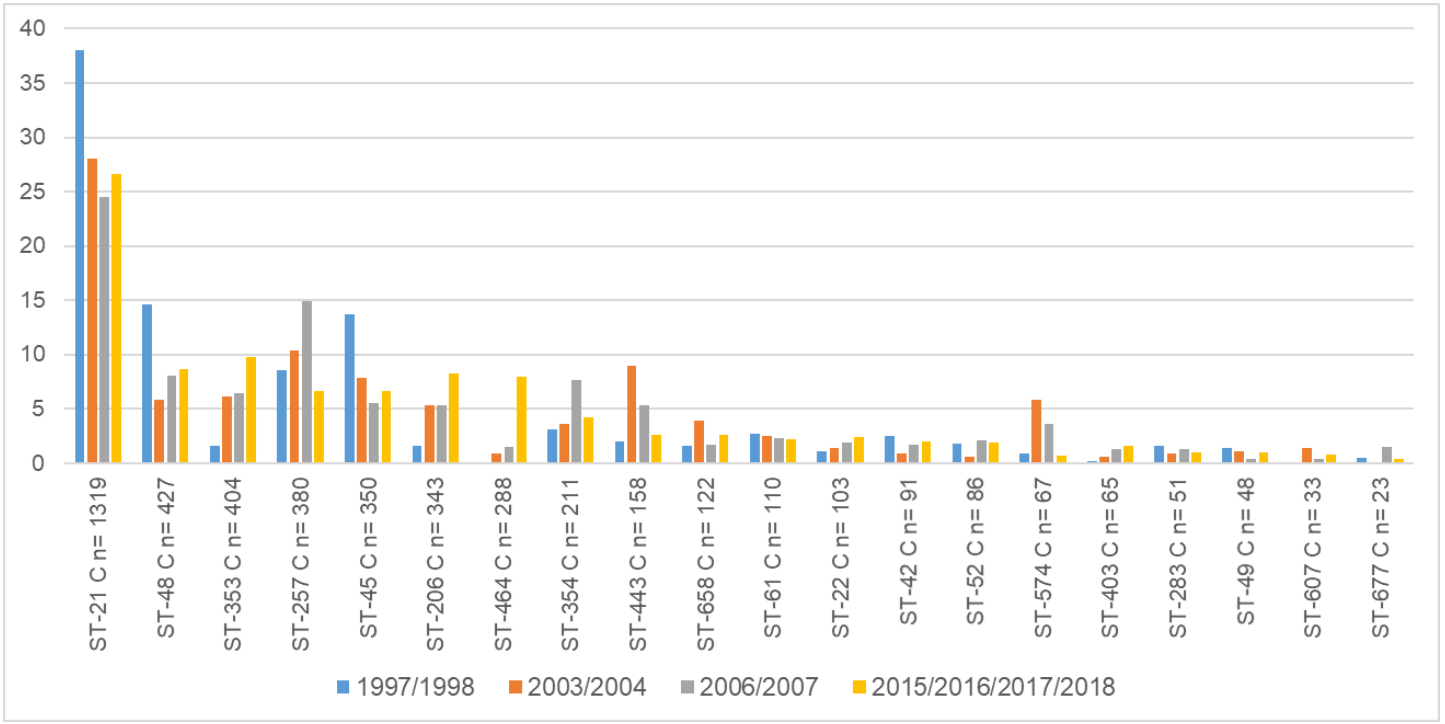



Figure 8: The distribution of each *C. jejuni* clonal complex, with total n>20 isolates, across 1997/8, 2003/4, 2005/6 and 2015/8, expressed as the percentage of all isolates that are of each clonal complex in that period.

### 4.4 *C. coli analysis*

Of the 240 North East *C. coli* isolates, 227 (94.5%) were assigned to ST-828 complex, one to ST-1150 complex and 12 unassigned so that analysis by clonal complex is not meaningful. Similarly, of the 151 *C. coli* from Oxfordshire samples 138 (91.3%) belonged to ST-828 complex and 13 were not assigned (Chi square p = 0.22). Considering sequence type, there is no statistical evidence for variation across the three years of the project (Chi square p=0.39) in the North East or in Oxfordshire, with the pattern among the more common types shown in Figure 9 and Figure 10 respectively.

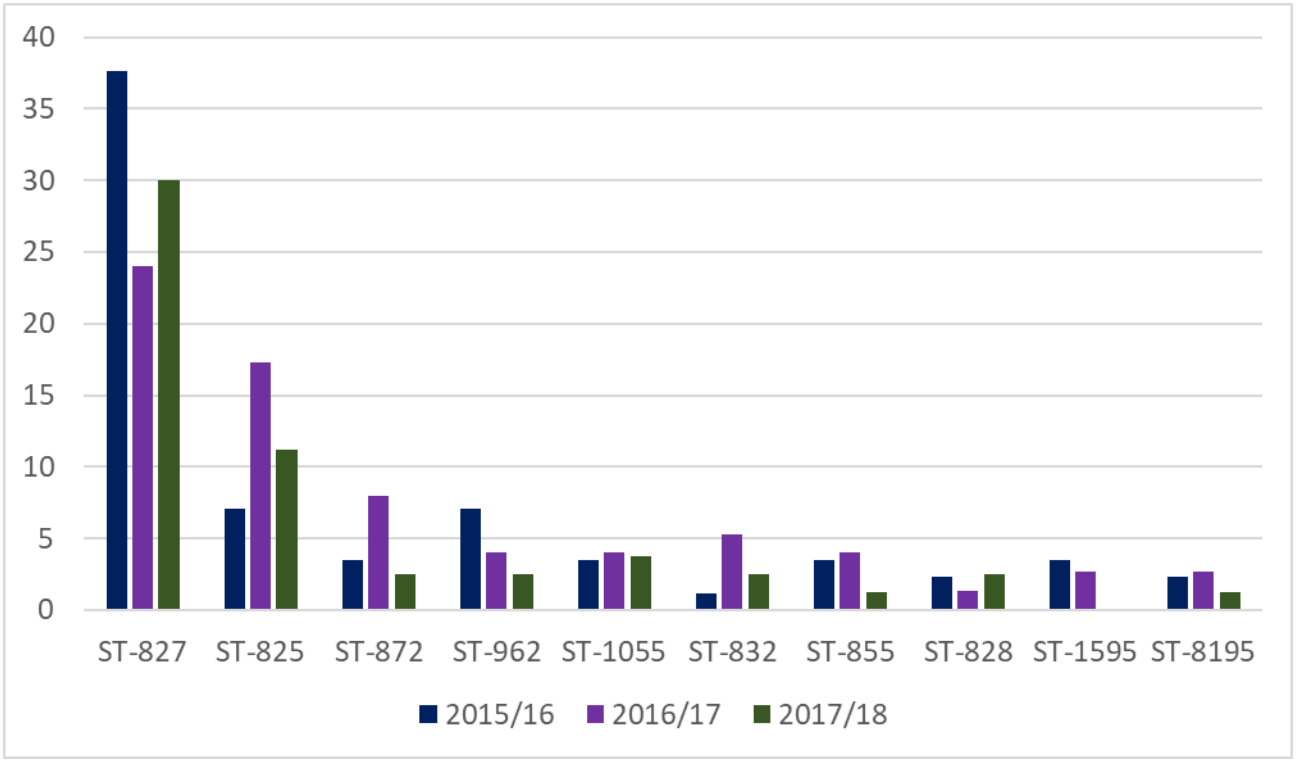



Figure 9. Percentage of *C. coli* isolates in the North East each year by sequence types for sequence types isolated at least 5 times across the study.

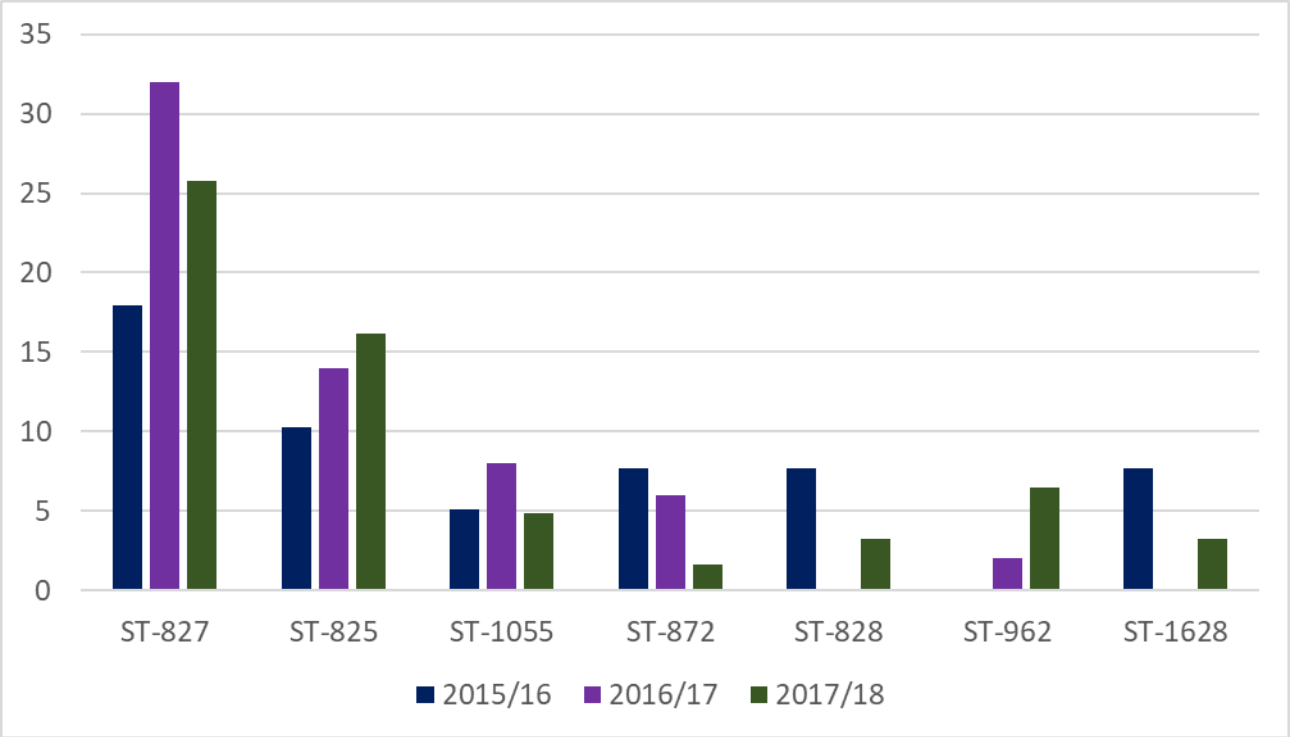



Figure 10. Percentage of *C. coli* isolates in Oxfordshire each year by sequence types for sequence types isolated at least 5 times across the study.

## 4.5 Conclusion

The distribution of clonal complexes was relatively stable across sites and over time in the current prospective sentinel surveillance study supporting the capacity of a sentinel surveillance approach to represent human infections nationally at this level of analysis. The pattern of distribution across sites and over time of tight genomic lineages in considered in analysis of clustering in Chapter 9. There was greater, and statistically significant evidence of variation over the longer time period from 1997/8 to 2015-8. However even over this much longer time scale the most prevalent clonal complexes in 1997/8 remain among the most prevalent in 2015-8, and there is substantial stability over time at this level of analysis. This shows that the same broad populations of detectable *Campylobacter* infecting residents in England persist over time as with some changes.

Using temporally and geographically distant reference isolates can potentially introduce bias in attribution (53). However population genetic algorithms also perform better with larger reference datasets (54). The relatively limited rate of change over time or across regional geographies in our dataset here, along with other work showing that host association across

non-human animals is stronger than international geographic variation (55) supports the use of relatively extensive host source reference datasets for source attribution analysis rather than precisely contemporaneous and sympatric, but much smaller, collections.

## 5. *Campylobacter* detection, enumeration, and characterisation from foods

There is very little available data on the extent of contamination on raw foods of animal origin at retail sale in England other than chicken meat. This section describes work to address this evidence gap. This work, sampling turkey, duck, and calf/ox livers and lamb livers complemented the contemporaneous FSA funded retail poultry survey. Quantitative microbiology was used to estimate the extent of contamination and whole genome sequencing to characterise the contaminating *Campylobacter* populations in each food type. This included testing samples by direct plating rather than enrichment to represent as closely as possible actual potential for human exposure.

### 5.1 Methods

The survey protocol used for the time-period from January 2017 to January 2019 is briefly described below.

#### Sampling

Sampling was undertaken from retailers in and around three widely separated laboratory centres from which the PHE food sampling network operates: York, London and Salisbury. Only fresh raw samples were tested and at the time of sampling, all were collected chilled. Samples for the survey were collected by trained individuals, who purchased samples from retail outlets and transported them to the testing laboratory according to the survey protocol. On arrival at the laboratory, the air temperature of the cool boxes was taken using calibrated data loggers or temperature probes. Sample details were logged onto the laboratory information management system.

#### Quantitative microbiology methods

All laboratories enumerated campylobacters based on EN/ISO 10272-2 for the enumeration of *Campylobacter* spp. using modified Charcoal Cefoperazone Deoxycholate Agar (mCCDA) as the primary plating medium. All participating laboratories used the same method of achieving a microaerophilic atmosphere (Campygen sachets, Thermofisher Ltd.). In brief, each sample was removed from its wrapping. Using sterile instruments (e.g. scissors and

tweezers) a 25 g of sample was aseptically removed and placed into a sterile bag. Buffered peptone water (BPW) was added so that a ratio of 1 part sample to 9 parts BPW weight was achieved and this was then homogenised for one minute. Portions of this sample were removed for enumeration by the surface plate method (ISO 10272-2:2017 Microbiology of food and animal feeding stuffs -- Horizontal method for detection and enumeration of *Campylobacter* spp. -- Part 2. Colony-count technique), and entailed the following: plating of 1 ml of the diluted homogenate onto three mCCDA plates (using Oxoid CM739 with Oxoid selective supplement SR155)) and 100 µl onto duplicate mCCDA plates. One further 10-fold dilution was prepared in maximum recovery diluent and 100µl plated onto one CCDA plate. CCDA plates were incubated in a microaerophilic atmosphere at 41.5 ± 1°C for 44 ± 4 h. The detection limit was 10 colony forming units (cfu) of campylobacters per g sample.

### Quality Assurance

All involved laboratories participate in recognised External Quality Assurance schemes (see here) including the FSA funded scheme for enumeration of *Campylobacter* species. Each also operate comprehensive internal quality assurance schemes as part of the requirements of their accreditation to ISO 17025/2005 as assessed annually by the United Kingdom Accreditation Service (UKAS).

### Statistical Analysis

Cross tabulations were analysed by calculating Clopper-Pearson and exact 95% confidence intervals for the proportion in each cfu per gram category. Additionally, Pearson chi square test of association was used to test the null hypothesis of no association between *Campylobacter* contamination and food animal source or season. Fisher's exact test was used for individual comparisons when sample numbers were small. Negative-binomial regression was used to estimate the independent associations between contamination (counts of colony forming units per gram) and both animal type and clonal complex.

## 5.2 Results

In all 1,890 samples were collected. The very large majority of duck meat and liver, ox/calf/lamb livers and turkey meat in retail shops appeared to originate from registered slaughterhouse premises that distribute UK wide.

## Sample descriptors

The large majority of samples were obtained from major retailers, except for duck liver, mainly obtained from other shops such as butchers (

Table 17). Among ovine liver samples all but one were labelled as lamb's livers, and one as sheep liver. Samples were relatively well spread across seasons but with some over-sampling of duck liver samples in winter months and some under-sampling of lamb's liver samples in spring months (Figure 11).

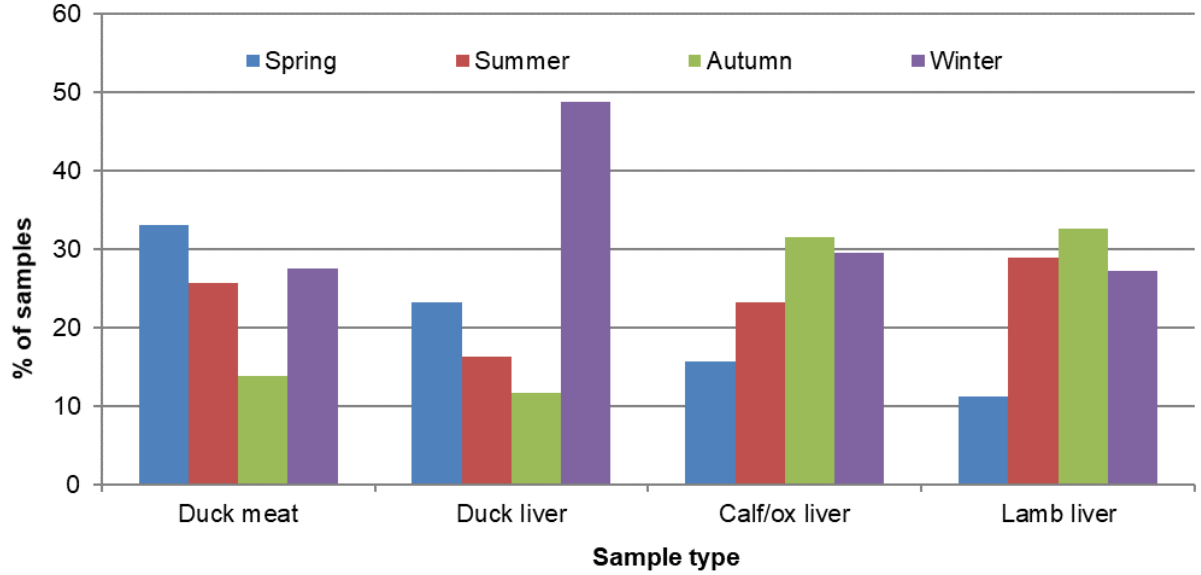



Figure 11. Proportion of samples tested by product type and season.

Table 17. Proportion of samples by food type from major and non-major retailers.

| Sample Type | Major Retailers (n) | Major Retailers (% of samples) | Non-Major retailers (n) | Non-Major retailers (% of samples) |
|---|---|---|---|---|
| Duck meat | 91 | 83 | 18 | 17 |
| Duck liver | 14 | 16 | 72 | 84 |
| Calf/Ox liver | 603 | 83 | 122 | 17 |
| Lamb/sheep liver | 618 | 84 | 122 | 16 |
| Turkey meat | 216 | 96 | 8 | 4 |
| Other | 4 | 67 | 2 | 33 |
| Total | 1546 | 82 | 344 | 18 |

* These were from one pig liver and five chicken liver samples that were also tested.

### Detection of *Campylobacter* spp. in samples

*Campylobacter* spp. were detected in 25.8% of samples and 1.4% (95% CI = 1% to 2%) of the samples had *Campylobacter* counts above 1,000 cfu per gram of sample. The highest single count detected was 57,000 cfu per g in a sample of lamb's liver.

### Numbers of *Campylobacter* spp. in relation to sample type

The proportion of turkey meat samples contaminated with *Campylobacter* spp. was significantly lower compared to any other sample group tested ($p < 0.001$; Fisher's exact test;

Table 18). Duck liver and duck meat samples were significantly more likely to be contaminated with campylobacters compared with the ruminant liver samples.
The proportion of samples with *Campylobacter* spp. levels at >1000 cfu per g ranged from 0 to 7 % across the sample types (

Table 18). Duck meat and duck liver samples were significantly more likely to have > 1,000 cfu per g compared to either ox/calf liver or lamb's liver samples (

Table 18). One pig liver sample was tested and found to contain 250 cfu of campylobacters per g. Five chicken liver samples tested were found to contain between 100 and 3,150 cfu of *Campylobacter* spp. per g.

Table 18. Comparison of contamination with *Campylobacter* spp. in retail meats and liver by animal species.

| **Sample type (n)** | **<10 (CI*)** | **10-1,000 (CI*)** | **> 1,000 (CI*)** |
|---|---|---|---|
| Duck meat (109) | 35 (26-45) | 61 (51-70) | 5 (2-10) |
| Duck liver (86) | 47 (36-58) | 47 (36-58) | 7 (3-16) |
| Calf/Ox liver (735) | 77 (74-90) | 22 (19-26) | 1 (0-2) |
| Lamb/sheep liver (740) | 74 (71-77) | 24 (21-28) | 1 (0-2) |
| Turkey meat (224) | 97 (94-99) | 3 (1-6) | 0 (0-1) |

*CI – Ninety-five percent confidence interval

## Detection of *Campylobacter* spp. in relation to season.

There was no significant difference in the proportion of samples with > 10 cfu of *Campylobacter* spp. per g among the different sampling seasons overall (Table 19) or for each sample type (Figure 12).

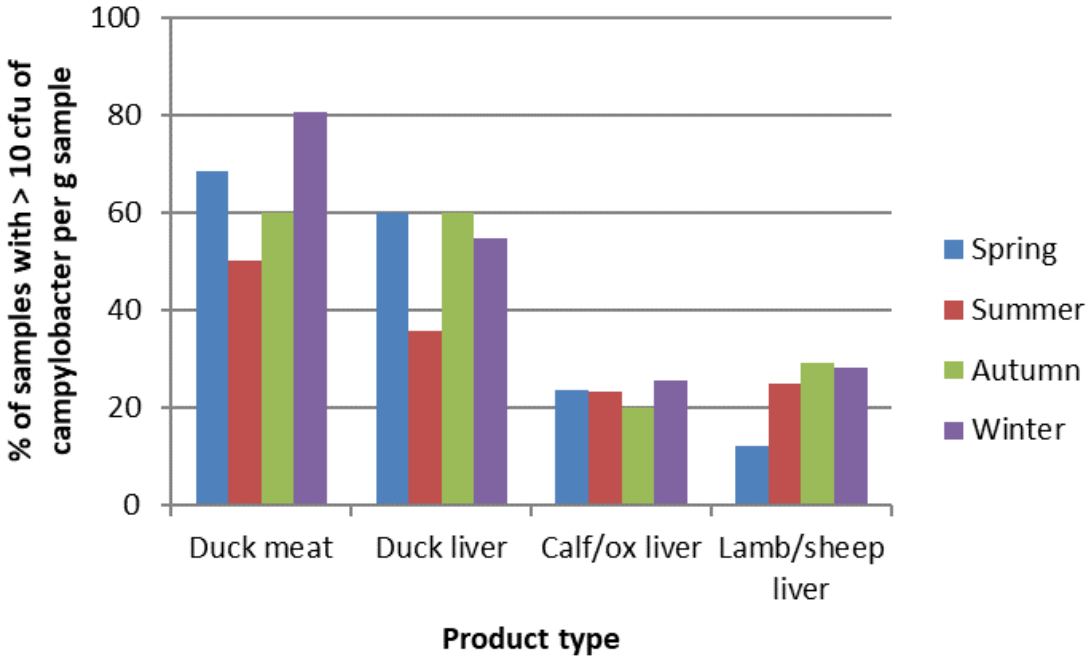



Figure 12. *Campylobacter* contamination in relation to season and product type.

There was no significant difference in the proportion of samples with > 1,000 cfu of *Campylobacter* spp. per g between the different sampling seasons (Table 19).

Table 19. Comparison of retail meat contamination with *Campylobacter* spp. by season as a percentage (%).

| **Season[a]** | **<10 (undetected) (CI[b])** | **10-1,000 (CI[b])** | **> 1,000 (CI[b])** |
|---|---|---|---|
| Spring | 75 (70-80) | 24 (18-28) | 1 (0.2-3) |
| Summer | 76 (72-80) | 23 (19-27) | 1 (0.3-2) |
| Autumn | 76 (72-79) | 23 (18-27) | 1 (0.1-2) |
| Winter | 70 (67-74_ | 27 (24-31) | 2 (1-4) |

[a]Spring months were March, April and May; summer June, July and August; autumn September, October and November; winter December, January and February.

[b] CI – Ninety-five percent confidence interval

## *Campylobacter* species isolated from samples

Isolates (440) from a total of 437 samples were subjected to *C. jejuni/C. coli* speciation testing. Of these *C. jejuni* alone was found in 83.9%, *C. coli* alone in 13.6%, both species in 2.3% of samples (Table 20). *C. lari* was detected in one duck liver sample. No speciation test or WGS was available for 50 isolates due to loss of isolate viability.

Table 20. *Campylobacter* spp. isolated from retail meat/liver samples.

| **Species detected** | **No. of isolates** | **% of isolates** |
|---|---|---|
| *C. jejuni* (only) | 369 | 83.9 |
| *C. coli* (only) | 60 | 13.6 |
| Mixed isolation (*C. jejuni* and *C. coli)* | 10 | 2.3 |
| *C. lari* | 1 | 0.2 |

Compared to *C. coli*, *C. jejuni* was significantly more likely to be isolated from lamb livers than from either duck or calf/ox liver samples (Fisher's exact test, $p < 0.001$, Table 21,

Table 22). For *C. jejuni* the relative abundance of clonal complexes is more similar across cattle and sheep than comparing either with duck origin isolates (Figure 13). The three isolates from turkey meat, four from chicken livers and one from pig liver were all speciated as *C. jejuni*. Chicken liver isolates were ST-257, ST-19, ST-5483 and ST-49; turkey isolates were ST-8555, ST-5 and ST-2844; the pig liver isolate was ST-556.

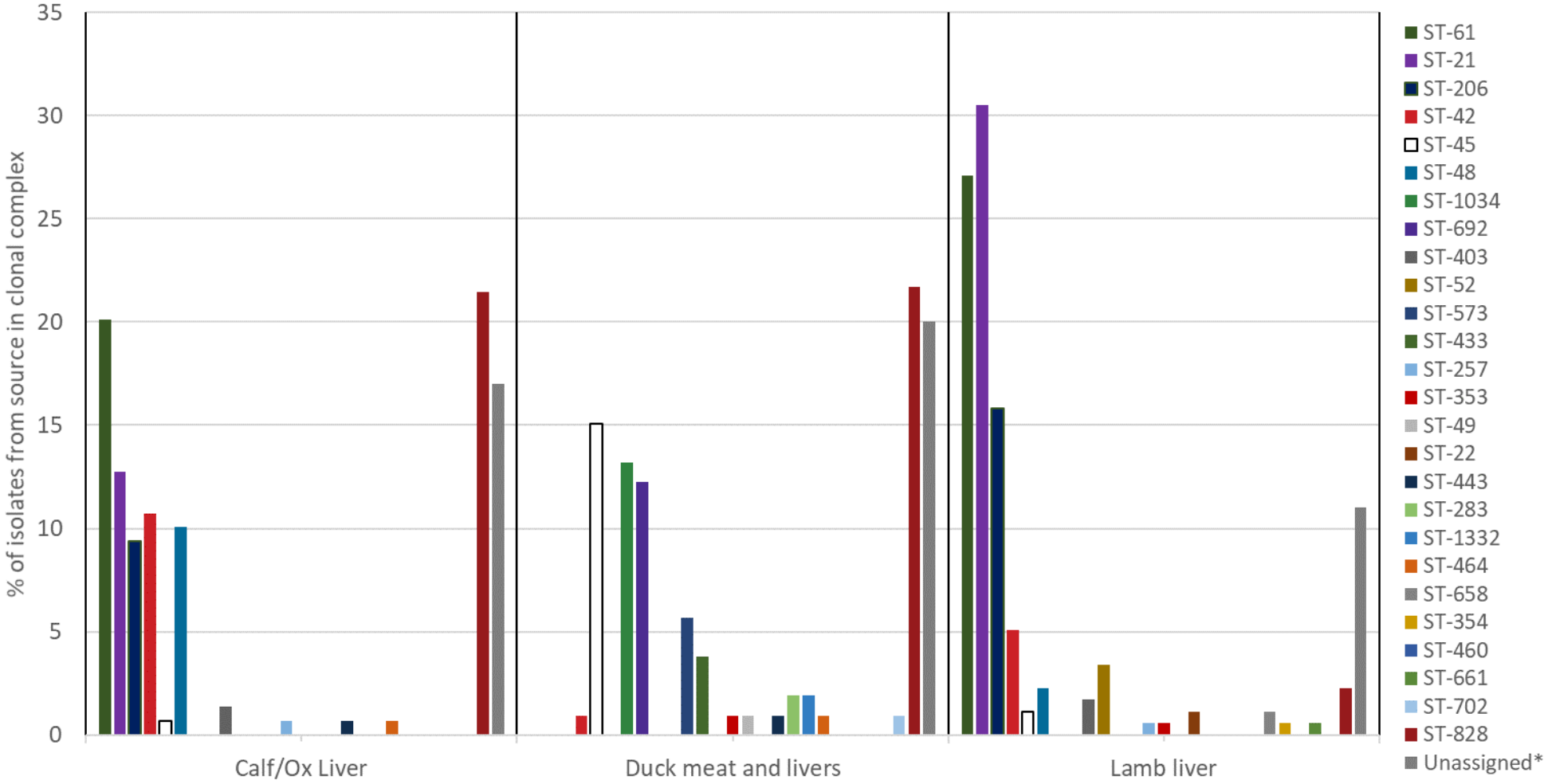



Figure 13. Distribution of clonal complexes across the main sample types (ST-828 complex is *C coli*, others are *C. jejuni*).

Table 21. Percentage of *C. jejuni* in each food animal species.

| Sample type | n | % | 95% CI |
|---|---|---|---|
| Duck meat or liver | 77 | 76.2 | 66.7-84.1 |
| Calf/Ox liver | 116 | 78.4 | 70.9-84.7 |
| Lamb/Sheep liver | 168 | 97.7 | 94.2-99.4 |

Table 22. Percentage of *C. coli* in each food animal species.

| Sample type | n | % | 95% CI |
|---|---|---|---|
| Duck meat or liver | 32 | 23.8 | 15.9-33.3 |
| Calf/Ox liver | 24 | 21.6 | 15.3-29.1 |
| Lamb/Sheep liver | 4 | 2.3 | 0.6-5.9 |

### Level of contamination in different foods and across clonal complexes

Analysis using a negative binomial regression model showed independent associations between cfu/g and both food source and clonal complex (Table 23, Table 24, Table 25). Exponentiated results in the table represent the ratios between counts compared to the reference category. Duck liver had higher counts than other samples and using ST-21 complex as a reference ST-48 complex, on average, had lower counts and ST-353 and ST-61 complexes higher.

## 5.3 Discussion

### Level of contamination compared to other foods

In this survey data set, *Campylobacter* spp. was detected in 60% of duck meat and liver, 26% of lamb or sheep liver, and 23% of calf/ox liver but just 3% of turkey samples. Overall 1.3% of samples had >1,000 cfu per gram of sample. When present, counts of *Campylobacter* were generally higher in duck liver than other samples tested. Counts also varied across among clonal complexes. This is the first report on the level of *Campylobacter*

contamination for these sample types at retail sale in England. Literature regarding contamination, and particularly quantified levels of contamination, is limited on meats at retail sale other than chicken. Past work in the UK has shown high levels of contamination in ruminant liver (42) while levels were rather lower in other meats of ruminant origin in Ireland (56) and more recently in Italy. This Italian study identified only a single isolate with a count above 10 cfu per gram in 1,203 bovine meat samples (57) while earlier studies did not quantify the level of contamination. The FSA funded retail poultry survey, reported separately to the FSA, shows higher levels of contamination, on average among poultry than the meat products from the food animals in this study.

### Comparison of types compared to other UK foods

The commonest clonal complexes for isolates originating from duck were ST-828, ST-45, ST-1034, ST-692 and ST-573 complexes, and from ox/calf livers were ST-828, ST-61, ST-21, ST-42, ST-48 and ST-206 complexes. In previous reports, ST-21, ST-42, ST-48 and ST-61 complexes were also associated with beef livers/cattle reservoir (58-61). In contrast to some previous studies, we detected relatively few ST-45 complex isolates but ST-206 was more common.

Table 23. Association of food type with counts of bacteria per gram in food samples.

| **Food type** | **Sample size (n)** | **Count ratio (95%CI)** | **p value** |
|---|---|---|---|
| Duck Liver | 23 | 1 (Reference) | |
| Duck Meat | 53 | 0.352 (0.153, 0.806) | **0.014** |
| Ruminant | 281 | 0.578 (0.281, 1.188) | 0.136 |

Table 24. Association of clonal complex with counts of bacteria per gram in food samples.

| **Clonal Complex** | **Sample size (n)** | **Count ratio (95%CI)** | **p value** |
|---|---|---|---|
| ST-21 complex | 78 | 1 (Reference) | |
| ST-61 complex | 73 | 1.999 (1.273, 3.14) | **0.003** |
| ST-257 complex | 32 | 0.236 (0.032, 1.74) | 0.157 |
| ST-353 complex | 26 | 184.358 (25.327, 1341.958) | **<0.001** |
| ST-206 complex | 19 | 1.285 (0.751, 2.199) | 0.360 |
| ST-464 complex | 19 | 0.162 (0.022, 1.199) | **0.075** |
| ST-45 complex | 14 | 3.321 (1.627, 6.781) | **<0.001** |
| ST-48 complex | 13 | 0.483 (0.236, 0.988) | **0.046** |
| Other or unassigned* | 83 | - | - |

Table 25. Association of food type and clonal complex with counts of bacteria per gram in food samples.

| **Food type** | **Clonal Complex** | **Sample size (n)** | **Count ratio (95%CI)** | **p value** |
|---|---|---|---|---|
| Duck Liver | - | 23 | 1 (Reference) | - |
| Duck Meat | - | 53 | 0.324 (0.158, 0.666) | **0.002** |
| Ruminant | - | 281 | 0.187 (0.076, 0.46) | **<0.001** |
| - | ST-21 complex | 78 | 1 (Reference) | **-** |
| - | ST-61 complex | 73 | 1.999 (1.281, 3.119) | **0.002** |
| - | ST-257 complex | 32 | 0.236 (0.033, 1.689) | 0.150 |
| - | ST-353 complex | 26 | 184.287 (25.028, 1356.93) | <0.001 |
| - | ST-206 complex | 19 | 1.285 (0.757, 2.181) | 0.353 |
| - | ST-464 complex | 19 | 0.147 (0.02, 1.104) | 0.062 |
| - | ST-45 complex | 14 | 1.181 (0.455, 3.073) | 0.733 |
| - | ST-48 complex | 13 | 0.483 (0.239, 0.977) | 0.043 |
| - | Other or unassigned* | 83 | - | - |

*Models included each clonal complex and merged unassigned isolates to a single group and evidence for this model is shown. Results for clonal complexes represented by 10 or more isolates are shown.

## 6. Attribution of human disease to source

Identification of the sources of human infection can guide interventions to reduce human infection from specific sources and can contribute to monitoring and the evaluation of interventions. The attribution of human *Campylobacter* infection to source based on genetic data is increasingly common. This section describes approaches to, and the results of this attribution. A systematic review of MLST-based attribution studies evaluated the approaches used and summarised their findings in the published literature. These findings informed the approach taken in the current study, including highlighting the lack of estimation of bias and adjustment for this bias in the existing published literature. A review of available whole genome-based attribution approaches was used to identify any candidate approaches that use the fuller data that is accruing with whole genome sequencing rather than just MLST. Datasets collated to support population genetic attribution of human disease are described. Methodological work included (i) validation studies to refine MLST-based attribution including measurement of bias and adjustment for this, and (ii) the development of sentinel-type based attribution. The results of attribution using these approaches on the study data are reported and compared.

### **6.1** Systematic review of MLST-based source attribution of *Campylobacter*

A systematic review of all peer reviewed publications that used multi-locus sequence typing data to attribute human disease isolates to source (with searches up to 23 November 2017) identified 25 publications for inclusion. This work is now accessible as a paper in Eurosurveillance with fuller details (38) and a summary provided here.

Reported studies were conducted across a wide geography but restricted to wealthy industrialised countries. Data sampling and analytical approaches varied, with five different attribution algorithms used. Poultry, in particular chicken, was identified as the principal source of human

infection and ruminant (cattle or sheep) sources were consistently implicated in a substantial proportion of cases. Specifically, the percentage of human disease attributed to poultry was reported in the range of 44% to 77% in studies applying the Structure population genetic algorithm, 57% to 83% in analyses using the Asymmetric Island model, and 52% to 80% in those using other analytical approaches (Figure 14). Variation among study results was substantial (with I-squared values above 90% for studies using both Structure and Asymmetric Island models). This level of heterogeneity does not support the calculation of a summary estimate across studies. Although studies varied in their precise results, in the populations studied, and in aspects of sampling and analysis they consistently identified the importance of poultry as a main source of human infection.

Methodologically, validation such as the accuracy of self-attribution of isolates from known sources, was reported in just five of 25 publications. No publication reported adjustment for biases identified by this validation. These gaps in validation and adjustment highlight this as an area for methodological development to generate improved estimates in future MLST and genomic attribution studies. As a result, and in the absence of an available method to improve attribution by using more extensive (whole genome) data as summarised in the next section, the study team extended planned validation work, developed a system to adjust attribution estimates to correct for identified biases implemented to improve the reliability of MLST based attribution. This and the development of a complementary sentinel-type approach (see section 6.7) were prioritised over population genetic analysis of whole genome data. These approaches aim to provide a framework to inform the application of future attribution approaches using more extensive genomic data.

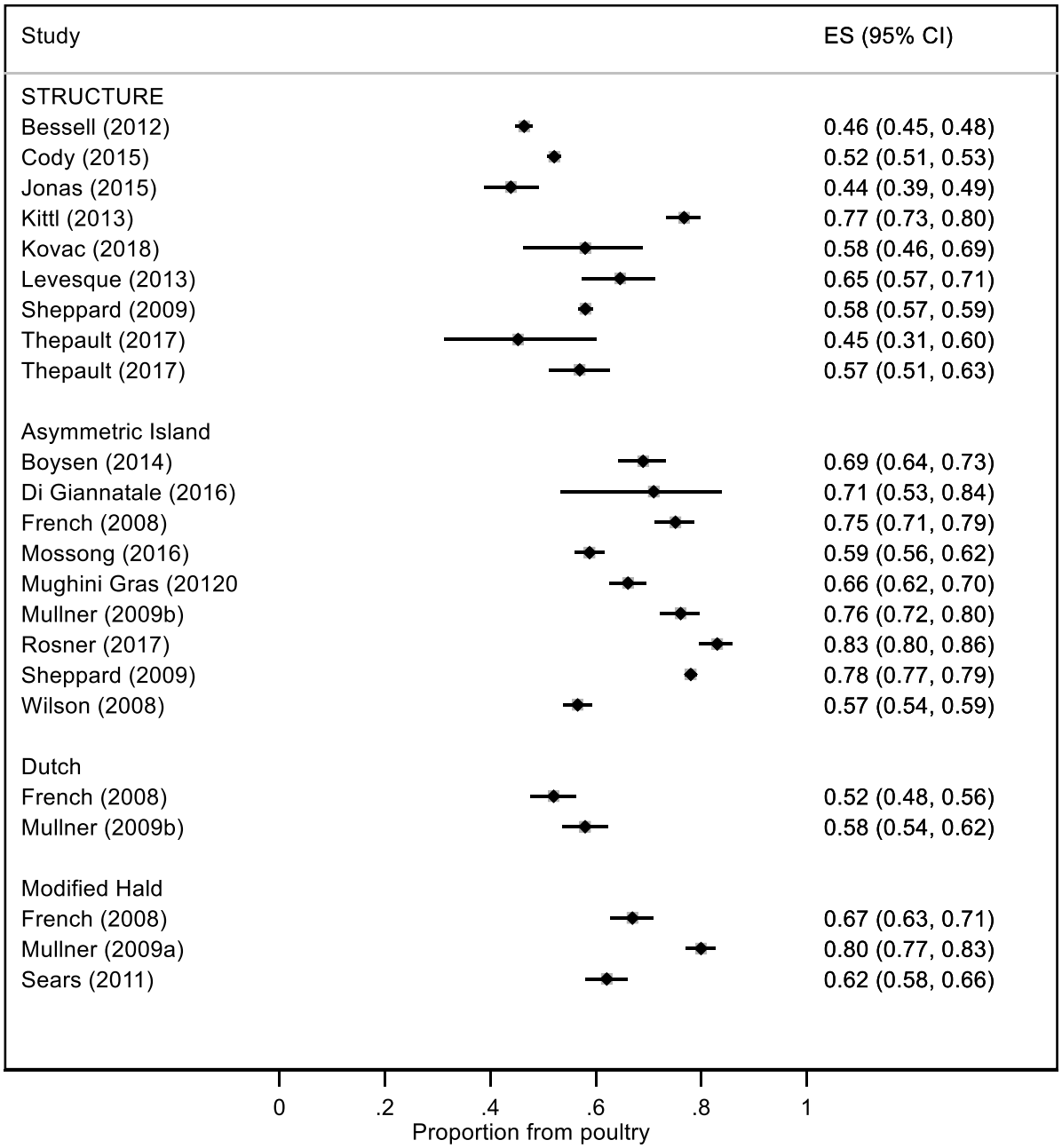

Study
ES (95% CI)
STRUCTURE
Bessell (2012)
0.46 (0.45, 0.48)
Cody (2015)
0.52 (0.51, 0.53)
Jonas (2015)
0.44 (0.39, 0.49)
Kittl (2013)
0.77 (0.73, 0.80)
Kovac (2018)
0.58 (0.46, 0.69)
Levesque (2013)
0.65 (0.57, 0.71)
Sheppard (2009)
0.58 (0.57, 0.59)
Thepault (2017)
0.45 (0.31, 0.60)
Thepault (2017)
0.57 (0.51, 0.63)
Asymmetric Island
Boysen (2014)
0.69 (0.64, 0.73)
Di Giannatale (2016)
0.71 (0.53, 0.84)
French (2008)
0.75 (0.71, 0.79)
Mossong (2016)
0.59 (0.56, 0.62)
Mughini Gras (20120
0.66 (0.62, 0.70)
Mullner (2009b)
0.76 (0.72, 0.80)
Rosner (2017)
0.83 (0.80, 0.86)
Sheppard (2009)
0.78 (0.77, 0.79)
Wilson (2008)
0.57 (0.54, 0.59)
Dutch
French (2008)
0.52 (0.48, 0.56)
Mullner (2009b)
0.58 (0.54, 0.62)
Modified Hald
French (2008)
0.67 (0.63, 0.71)
Mullner (2009a)
0.80 (0.77, 0.83)
Sears (2011)
0.62 (0.58, 0.66)
0
.2
.4
.6
.8
1
Proportion from poultry

Figure 14. Forest plot of the proportion of *C. jejuni* clinical isolates attributed to poultry by study, and uncertainty around these estimates. $I^2$ index > 90% for Structure and Asymmetric Island models so no summary estimates are given (Published in Eurosurveillance. Licence CC-BY). Confidence interval (95% CI) estimates are based on standard estimates for the precision of a proportion given the samples size rather than being true estimates of precision by each algorithm.

## 6.2 Whole genome sequence attribution

A review of published literature and update of this were performed in March 2016 and February 2018. No applicable published method for whole genome based attribution was identified. The review results are described more fully in Appendix 5. Follow up with authors identified that Prof Nigel French's team in New Zealand are developing a whole genome based implementation of the Asymmetric Island model. We have shared study data and this team including Prof French and Dr Jonathan Marshall at Massey University and Dr Simon Spencer at the University of Warwick, and are currently evaluating the approach on these data as ongoing work following completion of this main study reported here.

## 6.3 Reference datasets to support attribution

Datasets used in validation and attribution are summarised with fuller details in Appendix 6.

### MLST datasets for validation

Reference datasets were assembled from MLST profiles using the criteria that they were 1) on PubMLST, 2) from published studies, 3) from a source with at least 50 isolates, with the exception of including *C. jejuni* isolates from pig sources (n=26) given the potential importance of pig as a source of frequently consumed foods. Species assignment was confirmed, and publicly accessible projects saved on PubMLST as "*C. jejuni* attribution dataset" and "*C. coli* attribution dataset". Validation analysis was restricted to complete profiles. The numbers of isolates available by source and species using these criteria are given in

Table 26. The majority of *C. jejuni* isolates were from the UK, followed by Canada, Sweden, Switzerland and Luxembourg each with 500 or more and isolates from 26 other countries. The majority of *C. coli* isolates were also from the UK, followed by USA, Switzerland and Luxembourg each with 250 or more and isolates from 13 other countries. Included *C. jejuni* isolates ranged from 1981 to 2015 and *C. coli* from 1999 to 2015. Isolates in these datasets were therefore mainly geographically close to human populations being attributed to source using them reducing potential bias from this source. Human case isolates were generally later, and often substantially so, than those in the attribution dataset. As described in Chapter 4 lineage changes in human infections are relatively slow. Additionally, there isn't any published evidence of major changes in host association of the main lineages over time. These offer some reassurance on the use of non-contemporaneous isolates for attribution. However, we were not able to directly test the potential effect of bias given the relatively small numbers of contemporaneous source isolates.

Table 26. The number of isolates from *C. jejuni* and *C. coli* identified from sequence data used for attribution validation studies.

| **Source** | ***C. jejuni*** | ***C. coli*** |
|---|---|---|
| Cattle | 1,378 | 121 |
| Chicken | 5,011 | 2,430 |
| Dog | 213 | 0 |
| Duck | 129 | 129 |
| Environment | 624 | 168 |

| | | |
|---|---|---|
| Goose | 188 | 0 |
| Pig | 26 | 881 |
| Sheep | 390 | 89 |
| Turkey | 70 | 149 |
| Wild bird | 939 | 0 |
| Total | 8,968 | 3,967 |

Alongside new sampling and sequencing across meats at retail (turkey, duck, duck liver, lamb's liver, calf and ox liver) described in section 5, sequencing was undertaken on isolates from past chicken and pig abattoir surveys and the ongoing FSA retail chicken survey (FS102121) providing data to support development of whole genome sequence approaches to attribution (Table 27, Table 28). Datasets from all food animal sequencing accessible on PubMLST (Appendix 7).

Table 27. Number of *Campylobacter jejuni* and *Campylobacter coli* genomes sequenced from other FSA funded retail chicken and chicken and pig abattoir studies.

| **Source** | ***C. jejuni*** | ***C. coli*** | **Other*** | **Total** |
|---|---|---|---|---|
| Retail chicken | 362 | 97 | 3 | 462 |
| Abattoir chicken | 1,157 | 311 | - | 1,468 |
| Abattoir pig* | 9 | 79 | 10 | 96 |

*More than one species detected in some samples.

Table 28. Number of *Campylobacter jejuni* and *Campylobacter coli* genomes sequenced from other FSA funded retail chicken, chicken and pig studies. **

| **Source** | ***C. jejuni*** | ***C. coli*** | **Total** |
|---|---|---|---|
| Retail chicken | 395 | 104 | 499 |
| Abattoir chicken | 1,115 | 297 | 1,412 |
| Abattoir pig* | 9 | 69 | 78 |

**PubMLST data excludes those not meeting PHE quality criteria. It includes additional retail chicken survey isolates sequenced with funding from the FSA *Campylobacter* retail chicken survey (FS102121).

### Final MLST datasets for attribution

Self-attribution accuracy increased substantially when analysis was restricted to the major potential sources of human infection compared to analyses across much wider host ranges as outlined under validation below (section 6.4). These restricted datasets comprised isolates from poultry, ruminant and wild bird sources for *C. jejuni*, and chicken, ruminant and pig sources for *C. coli*. Details of how to access the data are given in Appendix 6.

*C. jejuni*

The restricted *C. jejuni* reference dataset contained 7,715 isolates; 5,011 (65.0%) sourced from chicken, 1,765 (22.9%) from ruminants combined (and excluding faecal isolates of presumed ruminant origin), and 939 (12.2%) from wild birds. Samples were obtained between 1981 and 2015

from North America (673), South America (9), Europe (6,583), Africa (76), Asia (251) and Oceania (78); 45 isolates were of unknown geographical origin. Of the 7,715 isolates 7,669 were linked to one or more of 79 publications via their PubMed id.

*C. coli*

The restricted *C. coli* reference dataset comprised 3,521 isolates of which 2,430 (69.0%) were of chicken origin, 210 (6.0%) were sourced from ruminants and 881 (25.0%) were from pigs. These were isolated between 1999 and 2015 from North America (500), Europe (2,826), Africa (19), Asia (1) and Oceania (151); 24 isolates were of unknown geographical origin. Of the 3,521 isolates 3,507 (99.6%) were linked in the database to one or more of 37 publications via their PubMed id.

## Whole genome data

An attribution reference dataset of WGS isolates was also compiled to facilitate development of WGS-based attribution. This was derived from searches of international databases in the International Nucleotide Sequence Database Collaboration, submissions to our own database PubMLST, and isolates from the current study. The International Nucleotide Sequence Database Collaboration comprises the European Nucleotide Archive (ENA), National Center for Biotechnology Information (NCBI) Genbank and the DNA DataBank of Japan (DDBJ). Each database has its own set of submission and retrieval tools but data is exchanged between them on a daily basis. Searches were on the ENA using the search terms '*Campylobacter jejuni* 'or '*Campylobacter coli*' were repeated at least annually. Resulting genome assembly contig sets were assessed for isolates suitable for inclusion in the PubMLST database, and were uploaded to the PubMLST database, if not already included, and linked via the PubMed id number to the original publication.

In constructing a dataset to support whole genome attribution assembled genomes belonging to coherent datasets were selected, rather than unpublished single isolate submissions with little or no useful metadata. The total number of genomes usable for attribution of human disease, with updates to 19 February 2019 are as outlined in

Table 29 and

Table 30. Details on how to access these genetic data are described in Appendix 6.

Table 29. Whole genome attribution dataset compiled from published literature, genome archives and pubmlst.org/campylobacter submissions, for *C. jejuni* isolates, as of 19.02.19, and available as genome assemblies at the pubmlst.org/campylobacter.

| **Source** | **Data source** | **Isolates** | **Continent*** |
|---|---|---|---|
| Chicken | Published | 519 | Europe, N. America |
| Chicken | PubMLST / ENA | 315 | Europe, Oceania, S. America, N. America |
| Chicken | APHA abattoir isolates | 1,115 | Europe |
| Chicken | FS101013 food isolates | 395 | Europe |
| Chicken | Food Standards Scotland | 1,176 | Europe |
| Chicken | Total | 3520 | - |
| Ruminant | Published | 103 | Europe , N. America |
| Ruminant | PubMLST / ENA | 152 | N. America, Oceania, Europe |
| Ruminant | FS101013 food isolates | 233 | Europe |
| Ruminant | Food Standards Scotland | 648 | Europe |
| Ruminant | Total | 1,136 | - |
| Wild bird | Published | 231 | Europe , N. America, Asia |
| Wild bird | PubMLST / ENA | 20 | Europe , N. America |

| | | | |
|---|---|---|---|
| Wild bird | Food Standards Scotland | 37 | Europe |
| Wild bird | Total | 288 | - |
| All | Grand Total | 4,944 | - |

*Continents are listed in order of isolate prevalence.

Table 30. Whole genome attribution dataset compiled from published literature, genome archives and pubmlst.org/campylobacter submissions, for *C. coli* isolates, as of 19.02.19, and available as genome assemblies at the pubmlst.org/campylobacter.

| **Source** | **Data source** | **Isolates** | **Continent*** |
|---|---|---|---|
| Chicken | Published | 282 | Europe, N. America, Asia |
| Chicken | PubMLST / ENA | 34 | N. America, Europe |
| Chicken | APHA abattoir isolates | 298 | Europe (all UK) |
| Chicken | FS101013 food isolates | 104 | Europe (all UK) |
| Chicken | Food Standards Scotland | 66 | Europe (all UK) |
| Chicken | Total | 784 | - |
| Ruminant | Published | 21 | N. America, Oceania, Europe |
| Ruminant | PubMLST / ENA | 18 | N. America, Oceania, Europe |
| Ruminant | FS101013 food isolates | 27 | Europe (all UK) |
| Ruminant | Food Standards Scotland | 179 | Europe (all UK) |
| Ruminant | Total | 245 | - |
| Pig | Published | 35 | Europe , N. America, Asia |
| Pig | PubMLST / ENA | 75 | N. America, Europe |
| Pig | Food Standards Scotland | 112 | Europe (all UK) |
| Pig | DEFRA/FSA/HPA/VMD/BPEX | 72 | Europe (all UK) |
| Pig | Total | 294 | - |
| All | Grand Total | 1,323 | - |

*Continents are listed in order of isolate prevalence.

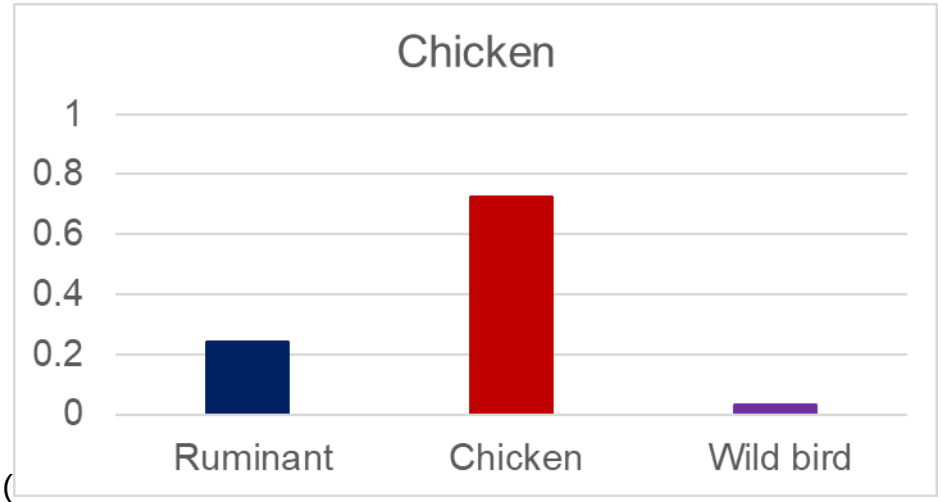
## 6.4 Validation of attribution using MLST data

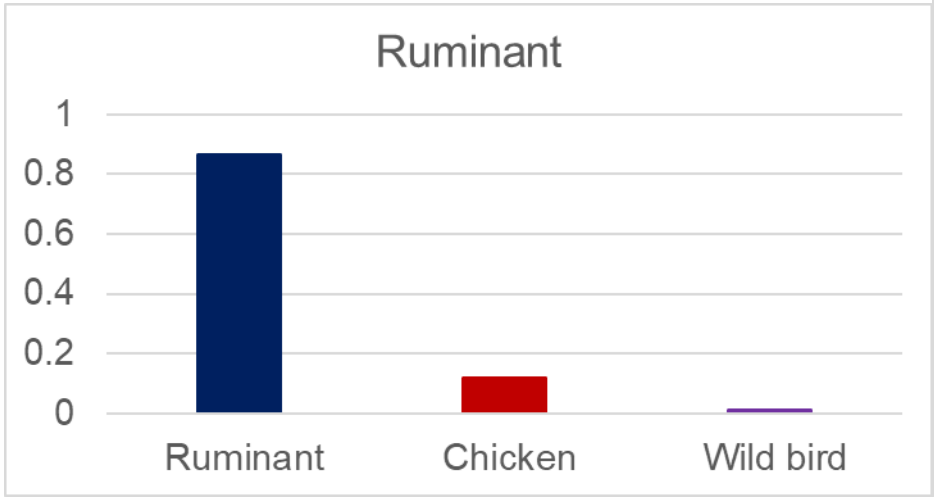
Structure (visited 18 August 2020) and iSource asymmetric island (visited 18 August 2020) models were applied to the datasets of known source. For Structure the no-admixture model was used, and 10,000 iterations followed 1,000 burn-in cycles. For iSource the asymmetric island model was run with a symmetric Dirichelet prior and 100,000 iterations following 10,000 burn-in cycles. Validation used self-attribution with 10% of the validation dataset having the known origin removed. These isolates with the true origin removed were attributed to sources by these population genetic algorithms based on the remaining isolates of known source in the validation dataset. This process was repeated 100 times and results averaged across these in each validation. Analyses evaluated the accuracy that could be achieved using broader datasets across a wide range of animals, and data restricted to the main sources considered important in human disease, and then proceeded to fuller comparison of algorithms.

### Assessment of the impact of using a wide and a narrow range of reservoir host species

Attribution using all available sources led to substantial inaccuracy (
Figure 15). For example dog and environment origin isolates were attributed broadly across wild and farm animal sources and sheep isolates attributed preferentially to cattle.

Removing minority and inaccurately assigned sources and merging of ruminant (cattle and sheep) sources led to improved accuracy for the main sources of interest (



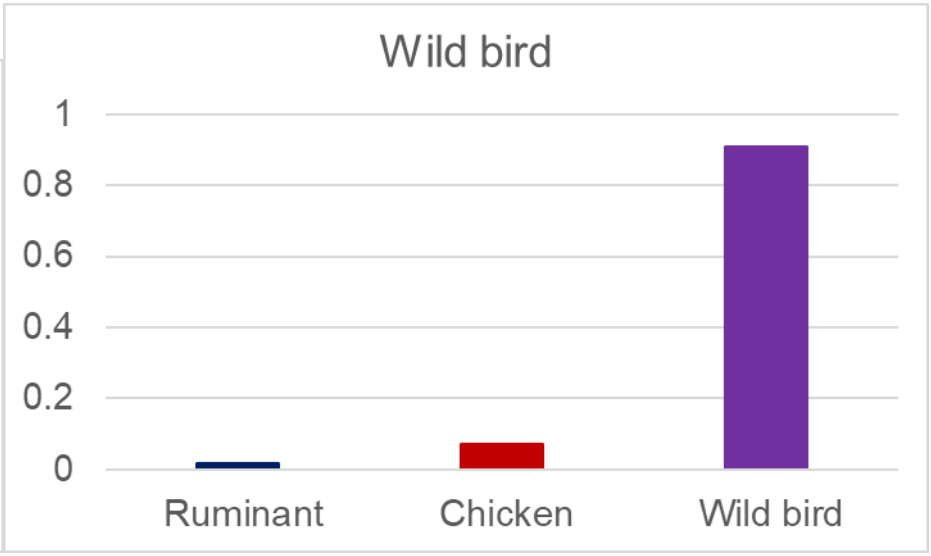



Figure 16) for *C. jejuni.* Similarly for *C. coli* attribution using the full range of available sources led to environmental isolates attributed across other sources, in particular to duck, sheep isolates miss-attributed substantially across chicken, pig and cattle, and inaccurate attribution between chicken and turkey. More accurate attribution was available by restriction to the sources of greatest interest and with the largest available data set (

Figure 17). Numbers per source are in Table 31

Similar inaccuracy was observed in attribution analyses using the iSource algorithm across a large range of possible sources with results summarised in confusion matrices (Figure 18. Proportional attribution to possible sources using iSource for *C. jejuni* (left panel) and *C. coli* (right panel) with true source along the bottom and colour coded attribution proportions to the possible sources listed along the vertical axis.

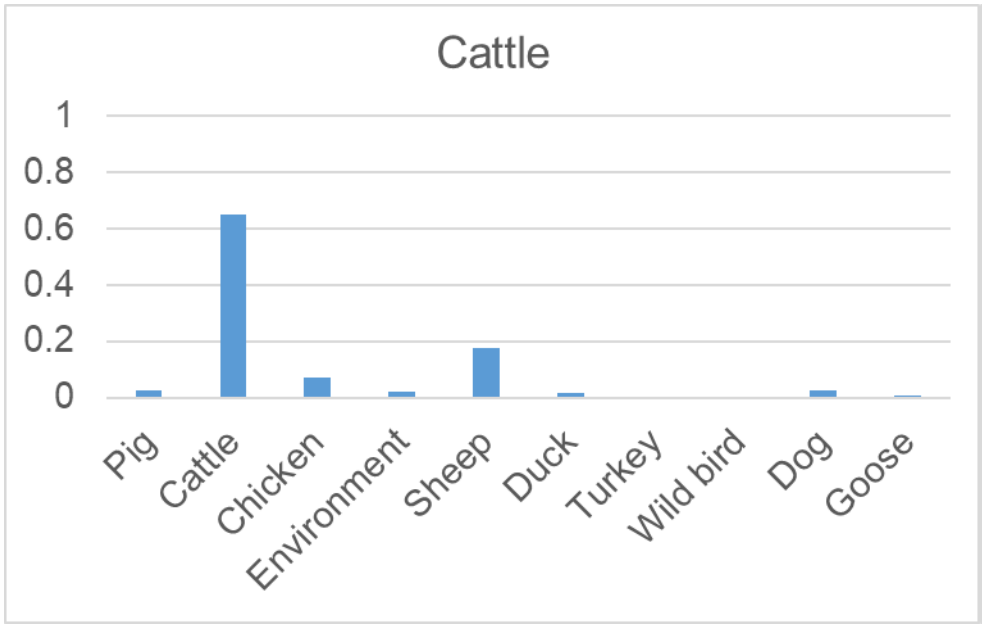
) showing the proportion assigned to each source (y-axis) in relation to the true source of the isolates (x-axis).

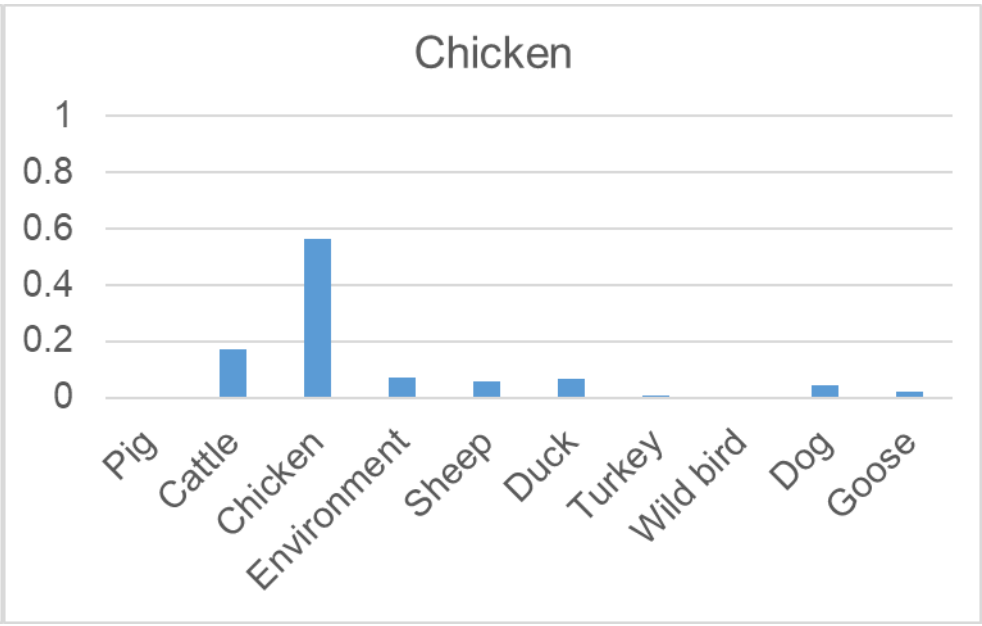
Table 31. Number of reference isolates from the main sources used in attributing human disease according to their species (*C. jejuni* and *C. coli*).

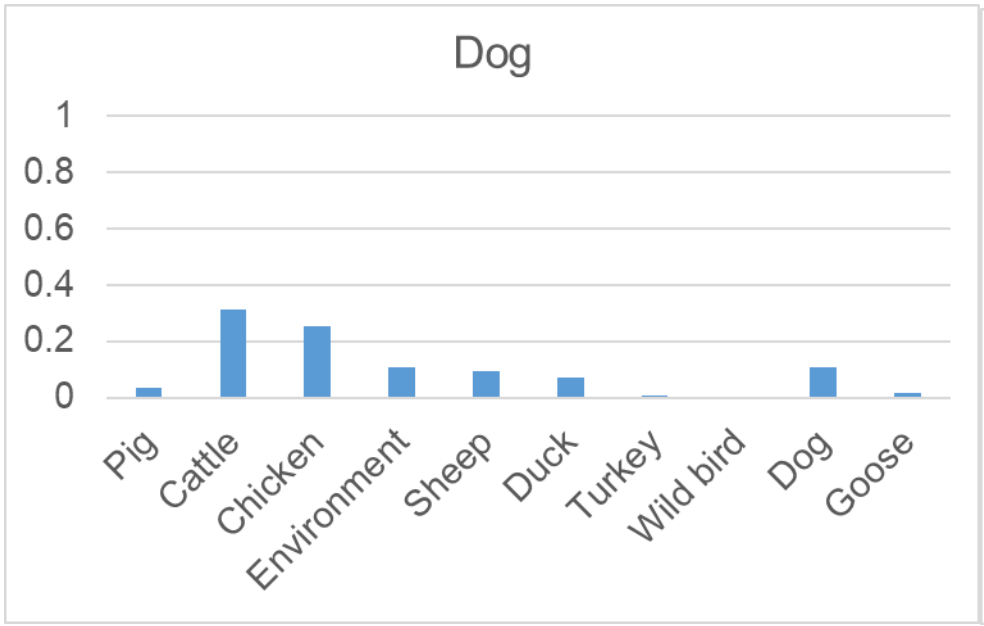
| **Source** | ***C. jejuni*** | ***C. coli*** |
|---|---|---|
| Chicken | 5,011 | 2,430 |
| Pig | 0 | 881 |
| Ruminant | 1,765* | 210 |
| Wild bird | 939 | 0 |
| Total | 7,715 | 3,521 |

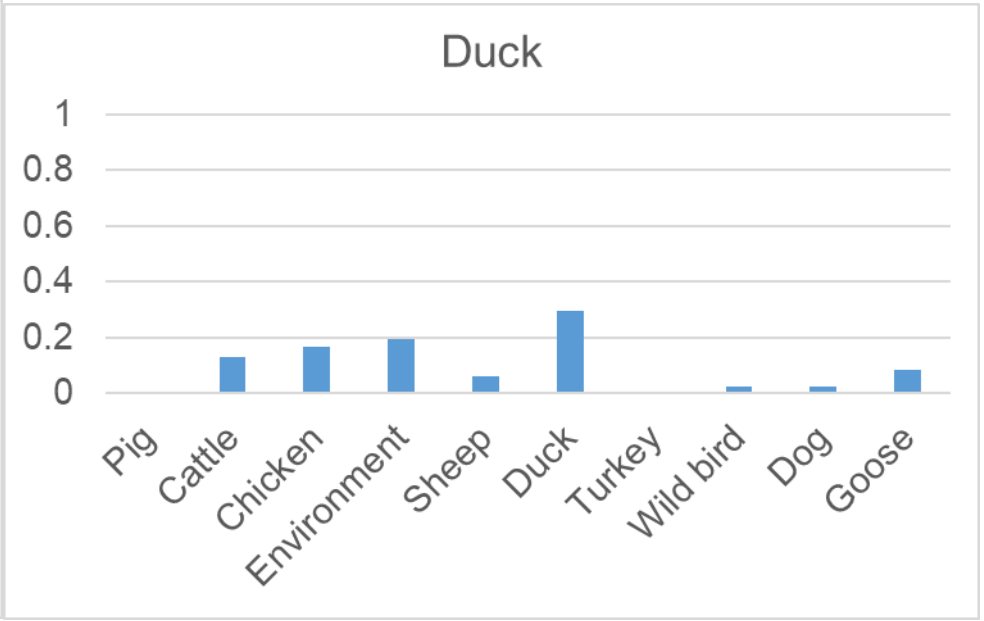
*Combining sheep and cattle isolates from

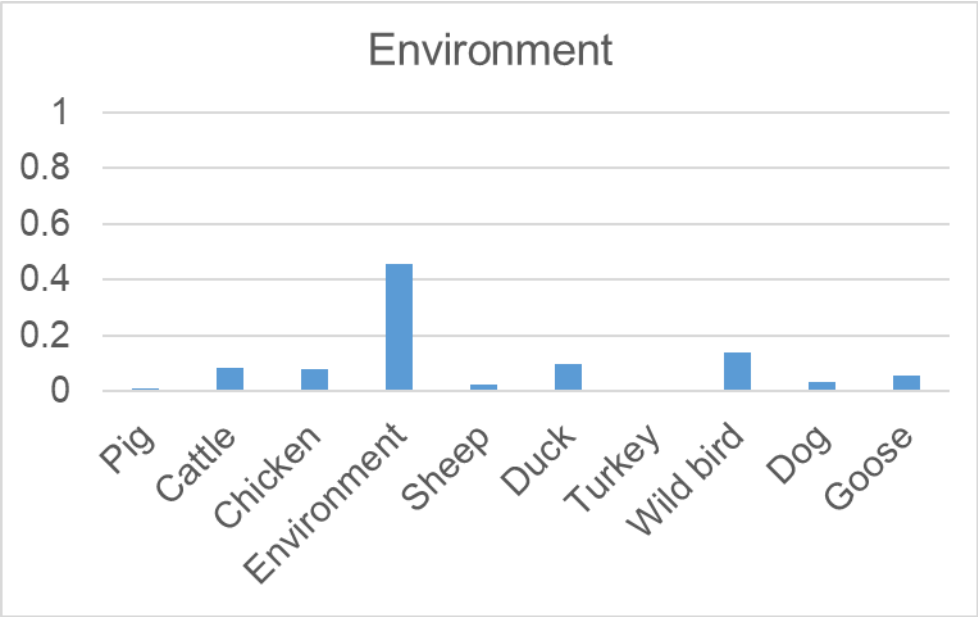
Table 26, removing isolates from faeces.

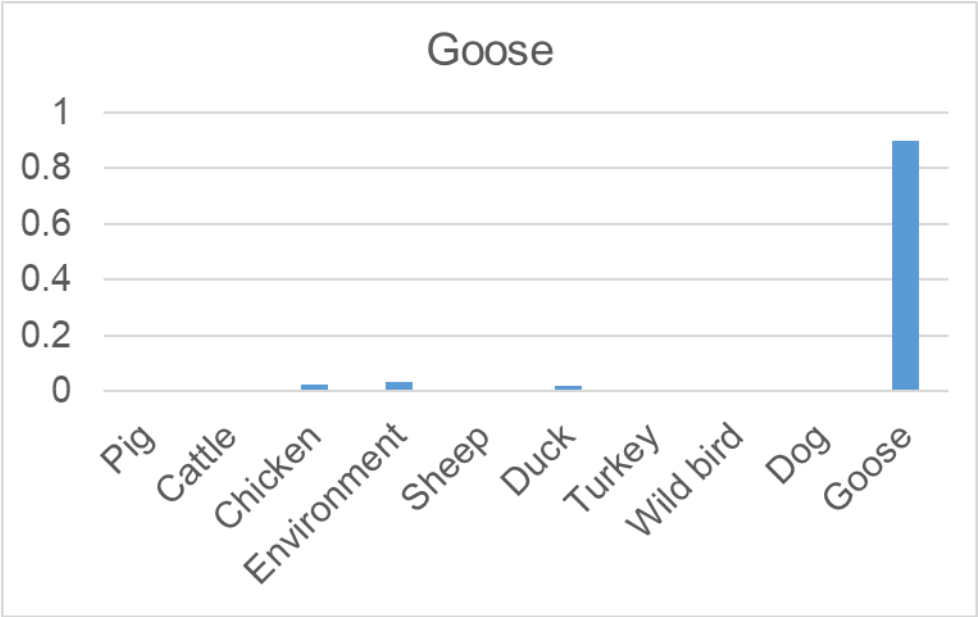

Cattle
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose

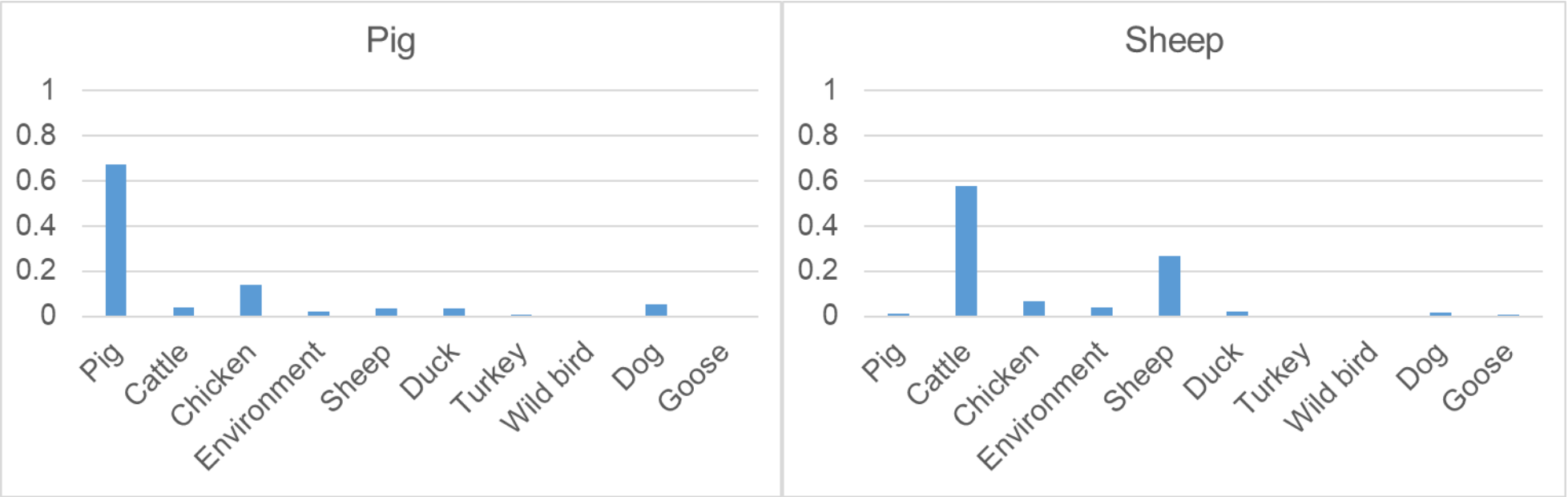

Chicken
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose


Dog
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose


Duck
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose


Environment
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose


Goose
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose


Pig
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose
Sheep
1
0.8
0.6
0.4
0.2
0
Pig
Cattle
Chicken
Environment
Sheep
Duck
Turkey
Wild bird
Dog
Goose

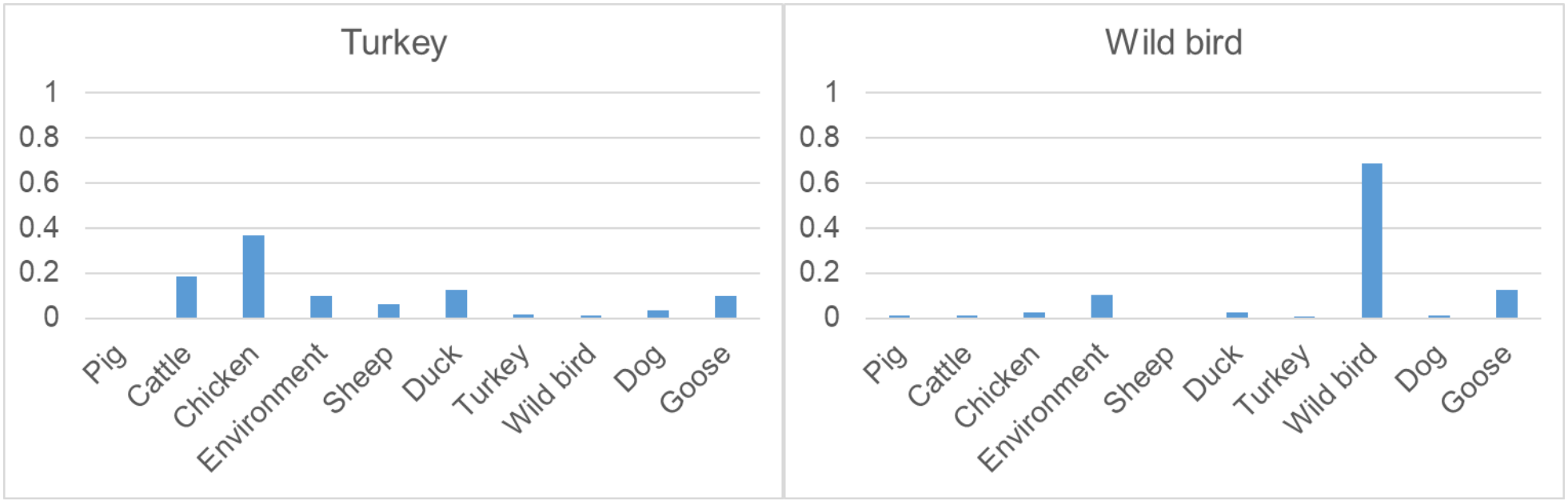
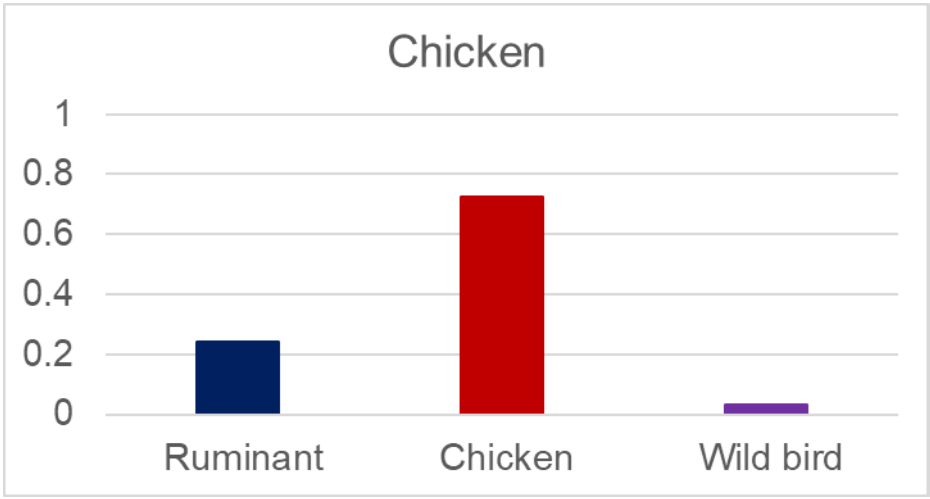
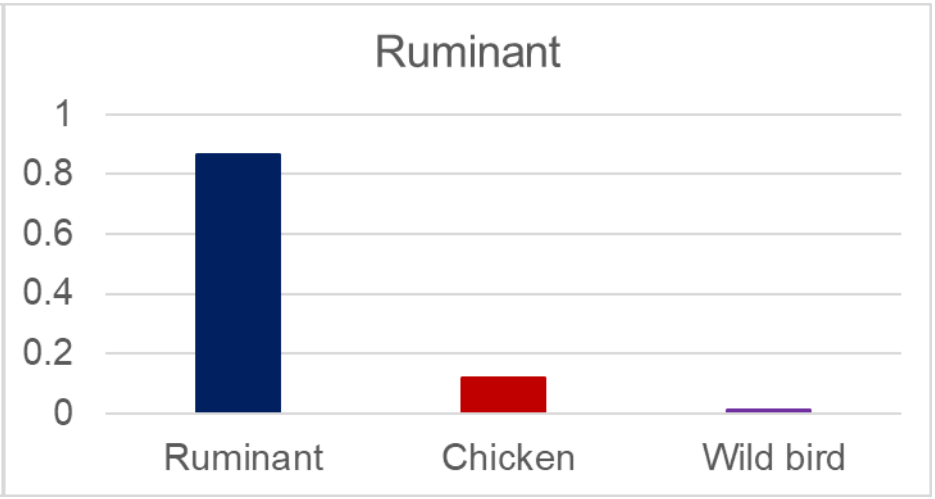



Figure 15. Proportional attribution of *C. jejuni* isolates to possible sources using Structure, with true origin indicated in each panel title.


Chicken
1
0.8
0.6
0.4
0.2
0
Ruminant
Chicken
Wild bird


Ruminant
1
0.8
0.6
0.4
0.2
0
Ruminant
Chicken
Wild bird

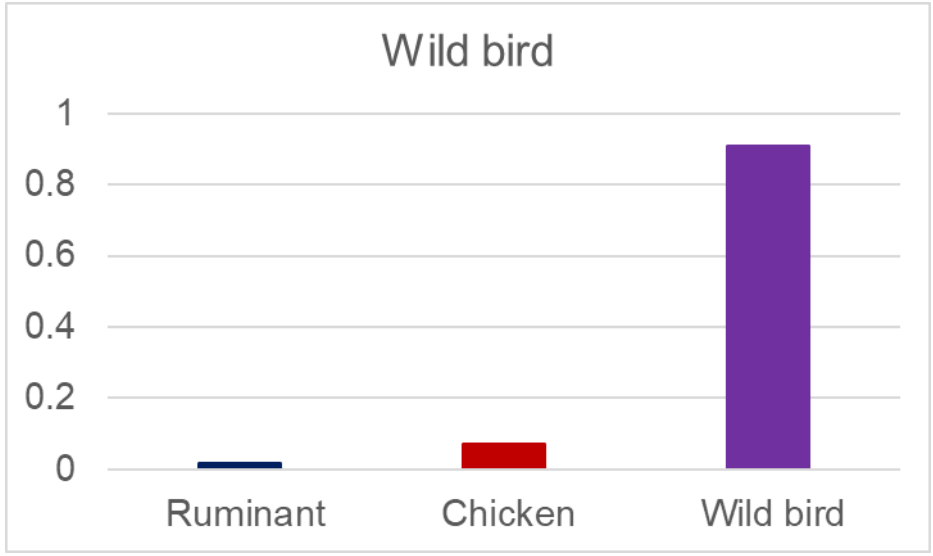
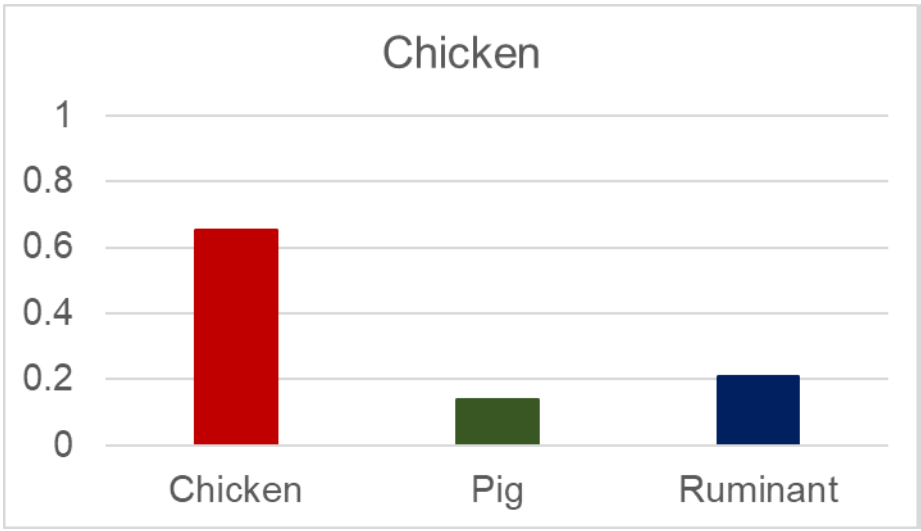
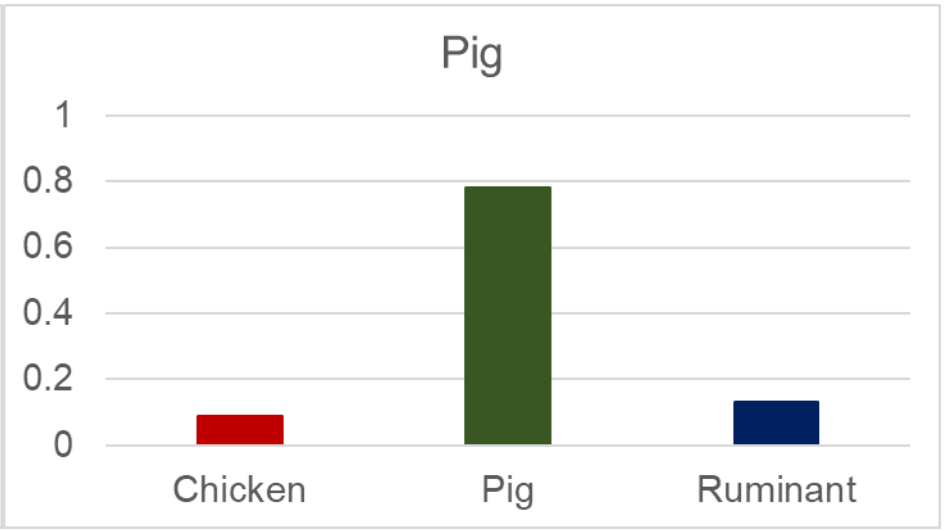

Wild bird
1
0.8
0.6
0.4
0.2
0
Ruminant
Chicken
Wild bird

Figure 16. Proportional attribution of *C. jejuni* isolates to possible sources using Structure in a dataset restricted to three sources, with true origin indicated in each panel title.


Chicken
1
0.8
0.6
0.4
0.2
0
Chicken
Pig
Ruminant


Pig
1
0.8
0.6
0.4
0.2
0
Chicken
Pig
Ruminant

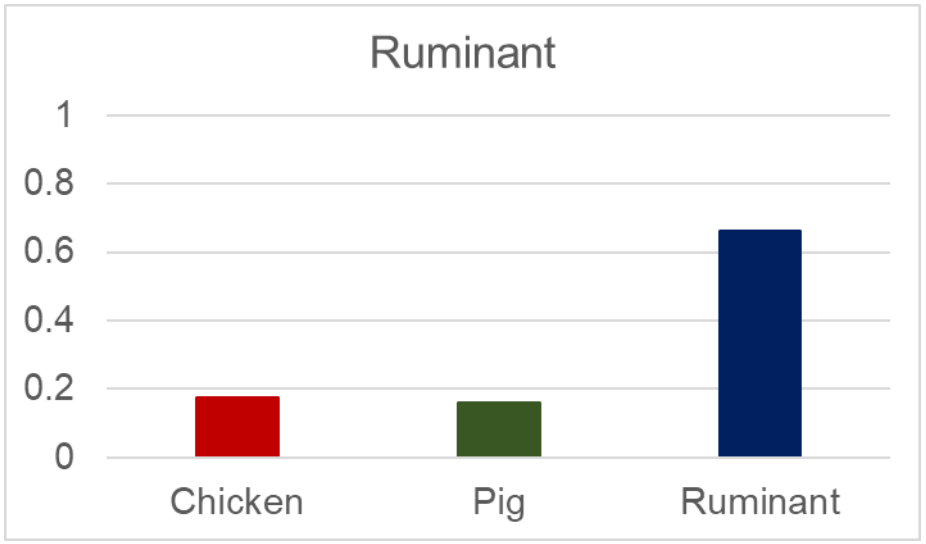

Ruminant
1
0.8
0.6
0.4
0.2
0
Chicken
Pig
Ruminant

Figure 17. Proportional attribution of *C. coli* isolates to possible sources using Structure in a dataset restricted to three sources, with true origin indicated in each panel title.

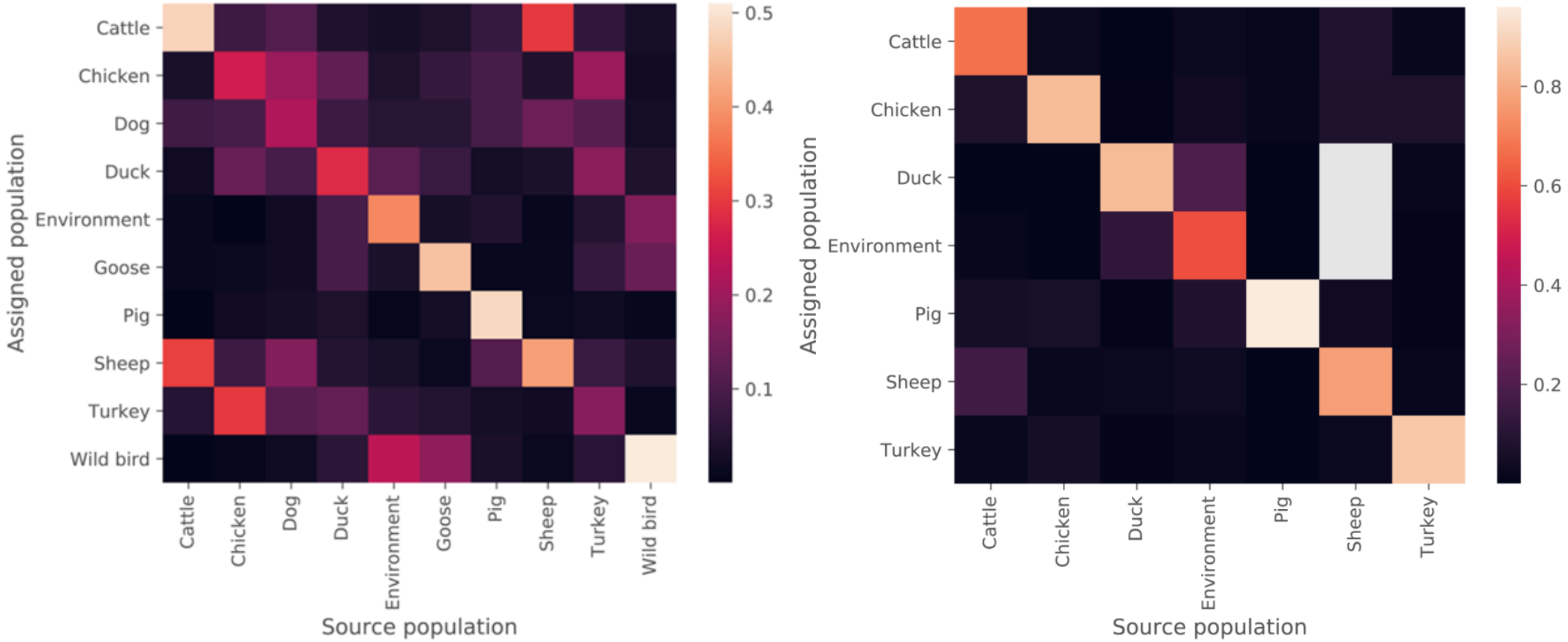



Figure 18. Proportional attribution to possible sources using iSource for *C. jejuni* (left panel) and *C. coli* (right panel) with true source along the bottom and colour coded attribution proportions to the possible sources listed along the vertical axis.

## Comparisons of algorithms

Using these three putative sources for each of *C. jejuni* and *C. coli* the algorithms were compared for accuracy using self-attribution as described above (6.4). iSource self-attributed isolates chicken origin to chicken more strongly but also strongly attributed *C. jejuni* isolates from ruminants and wild birds to chicken with substantially lower overall accuracy averaged across sources as a result of this, and similarly for ruminant and pig origin isolates for *C. coli.*(Figure 19,

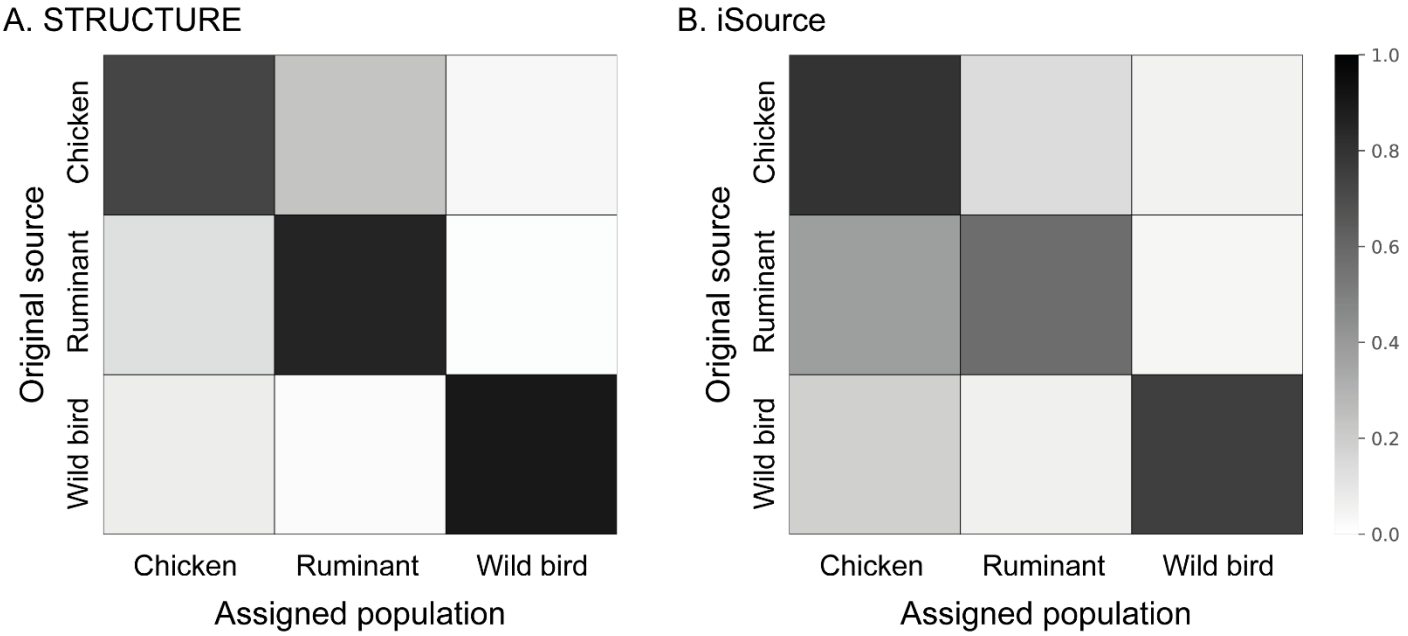
Table 33,

Table 35),.

Given the more consistent accuracy of Structure this was adopted as the main approach to attribution for the study. However, even using this approach there is substantial miss-attribution as shown in Figure 19  and Table 32,

Table 34. Considering, for example the Structure attribution of *C. jejuni* from known sources in Table 32: of isolates truly from chicken 73% are assigned to chicken while 27% are assigned to other sources (here mainly (24%) to ruminant). Moreover, these inaccuracies are not balanced. Of isolates from ruminants, only 13% were incorrectly assigned to chicken, substantially lower than the error in the opposite direction. The net effect is biased and would therefore also lead to biased attribution of human isolates. The direction and extent of the bias produced will depend on the actual sources of infections in the dataset being studied. This motivated work to correct for these biases, particularly in the light of no evidence in the peer reviewed published literature (section 6.1 above) for any correction of such biases identified in validation testing.

*Campylobacter jejuni*




*Campylobacter coli*

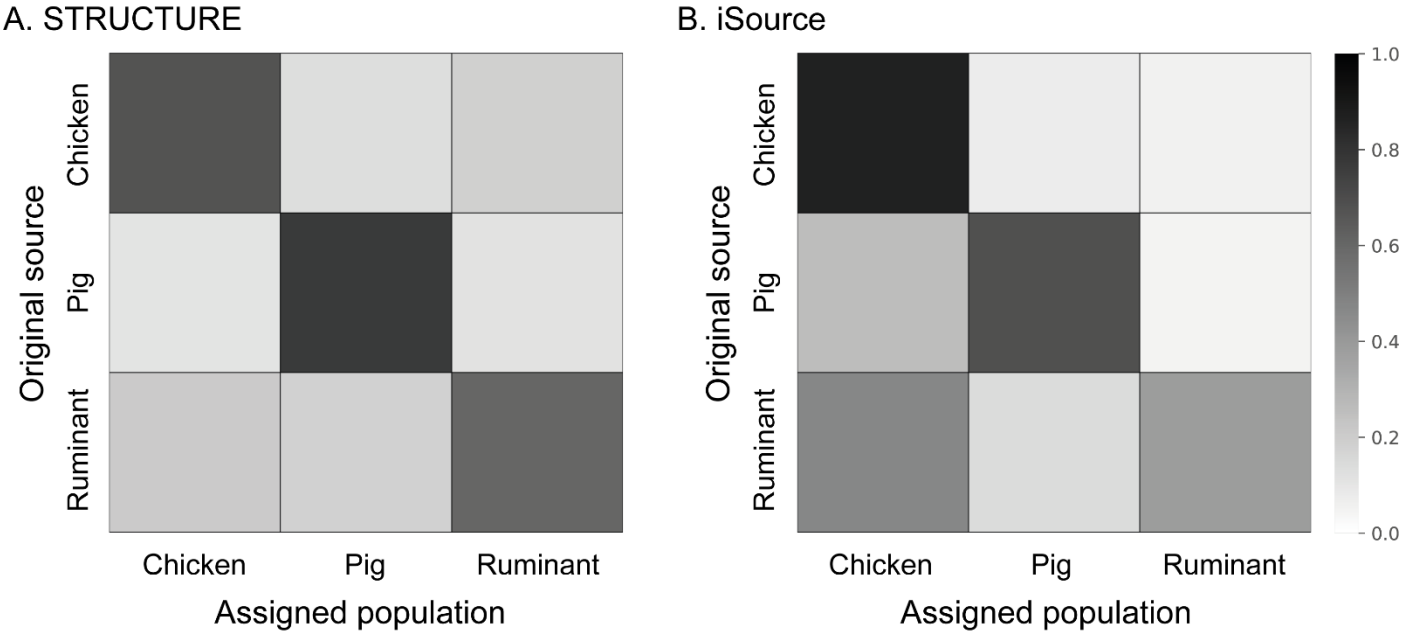



Figure 19. Proportional attribution to possible sources for *C. jejuni* (upper) and *C. coli* (lower panel) with true source along the side, assigned along the bottom and shading showing accuracy. Left hand panels (A) show Structure results and right (B) iSource. Numbers are given in

Table 32.

Table 32. Proportional attribution by Structure algorithms to possible sources for *C. jejuni* with true source along the side. Proportions are calculated by summing probabilities for source of origin in each source across the attributed isolates.

| **Source** | **Chicken** | **Ruminant** | **Wild bird** |
|---|---|---|---|
| **Chicken** | 0.73 | 0.24 | 0.03 |
| **Ruminant** | 0.13 | 0.87 | 0.01 |
| **Wild bird** | 0.07 | 0.02 | 0.91 |

Table 33. Proportional attribution by iSource algorithms to possible sources for C. jejuni with true source along the side. Proportions are calculated by summing probabilities for source of origin in each source across the attributed isolates.

| **Source** | **Chicken** | **Ruminant** | **Wild bird** |
|---|---|---|---|
| **Chicken** | 0.81 | 0.14 | 0.05 |
| **Ruminant** | 0.38 | 0.58 | 0.04 |
| **Wild bird** | 0.19 | 0.06 | 0.75 |

Table 34. Proportional attribution by Structure algorithms to possible sources for C. coli with true source along the side. Proportions are calculated by summing probabilities for source of origin in each source across the attributed isolates.

| **Source** | **Chicken** | **Pig** | **Ruminant** |
|---|---|---|---|
| **Chicken** | 0.68 | 0.14 | 0.19 |
| **Pig** | 0.11 | 0.78 | 0.12 |
| **Ruminant** | 0.21 | 0.18 | 0.61 |

Table 35. Proportional attribution by iSource algorithms to possible sources for C. coli with true source along the side. Proportions are calculated by summing probabilities for source of origin in each source across the attributed isolates.

| **Source** | **Chicken** | **Pig** | **Ruminant** |
|---|---|---|---|
| **Chicken** | 0.87 | 0.07 | 0.06 |
| **Pig** | 0.26 | 0.69 | 0.05 |
| **Ruminant** | 0.47 | 0.14 | 0.39 |

## 6.5 Adjustment for bias in MLST based attribution

Considering for example the results of attribution to source of *C. coli* isolates using the Structure algorithm reported in

Table 34: the isolates attributed to chicken comprise 68% of isolates from chicken, along with 11% of isolates actually from pig, and 21% of those actually from ruminants. Similarly, the totals attributed to pig and to ruminants are comprised of isolates that actually come from all three sources in proportions as given in

Table 34. Using these as estimates of error leading to bias it is possible to adjust for that bias in attributing human infections. The approach used to adjust for bias was to create and solve simultaneous equations, bounding solutions at zero for each source (since no source could contribute less than 0%) and constraining the totals to 100%. This allows calculation of the proportion of human disease from each source adjusted for the observed biases as estimated in the validation work. This adjustment assumes that the same biases apply when attributing isolates from humans, as well as restricting the possible origin of human infection to the sources being studied. It offers estimates of the underlying source and not any inference on the transmission route to humans as regards direct, foodborne, or environmental.

This is shown diagrammatically in Figure 20. The estimates obtained by running the attribution algorithm (shown in black) are related to the preferred but unknown true values from each source (x for chicken, y for pig and z for ruminant). Estimates for the proportion of isolates from chicken actually being assigned to chicken (0.68), from pig being mistakenly attributed to chicken (0.11) etc are taken from the validation study results (

Table 34) to fill in numbers in place of each combination picture of two animals (real origin in colour, attributed origin in grey) in parenthesis. This gives the equations as below with C the proportion of a human *C. coli* isolate dataset attributed to chicken, P the proportion to pig, and R to ruminant before this adjustment.

$0.68x + 0.11y + 0.21z = C$, $0.14x + 0.78y + 0.18z = P$, $019x + 0.12y + 0.61z = R$,

The proportions from each source, chicken (x), pig (y), and ruminant (z) are then estimated by solving the simultaneous equations having filled these numbers from the above tables and the raw results from the attribution analysis of human isolates into the places indicated by black animals.

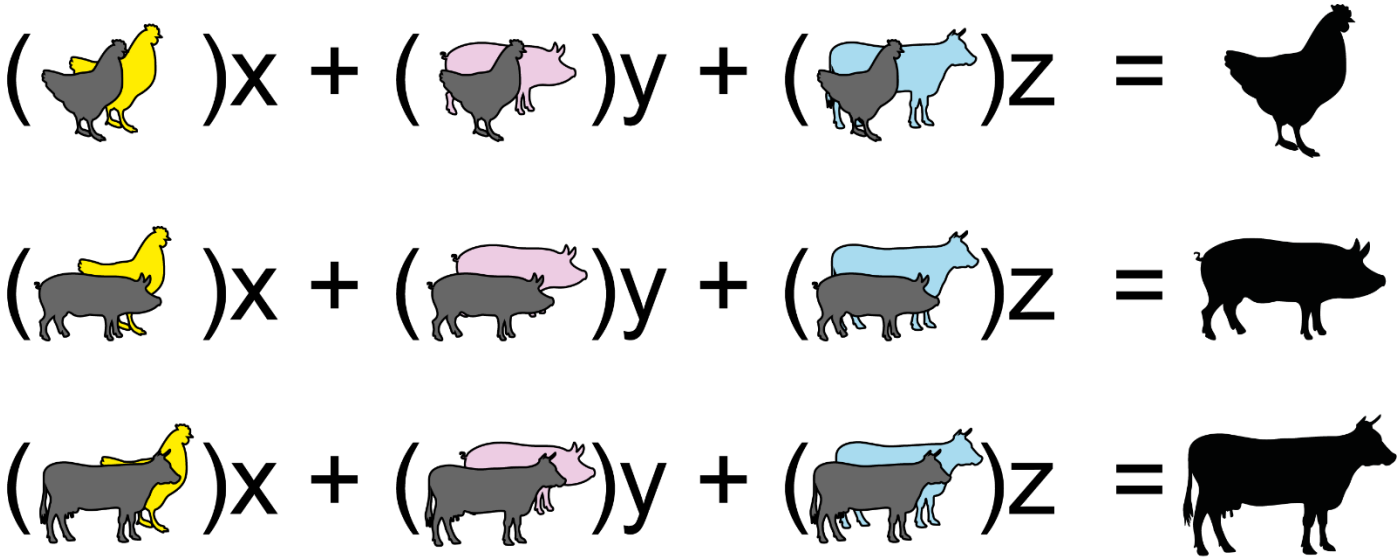


Figure 20. Graphical illustration of the approach to adjustment for bias estimated in validation. The estimates obtained by unadjusted attribution are in black, while what is wanted is the true proportion from each source, algebraically x for chicken, y for pig and z for ruminant. The proportion of isolates from each true source (coloured) animal to attributed to each animal species (greyed animal), are estimated from validation studies.

### 6.6 MLST based attribution results, raw and adjusted for identified bias

The following figures show attribution by study year for Oxfordshire from 2004 to 2018 including historical data, and for data from the current study by quarter from October 2015 to September 2018. Annual Oxfordshire attribution years run from October to September. The study year is allocated based on the later 9 months of this period in all graphs (October 2003 to September 2004 as 2004 etc). Data are tabulated in Appendix 8.

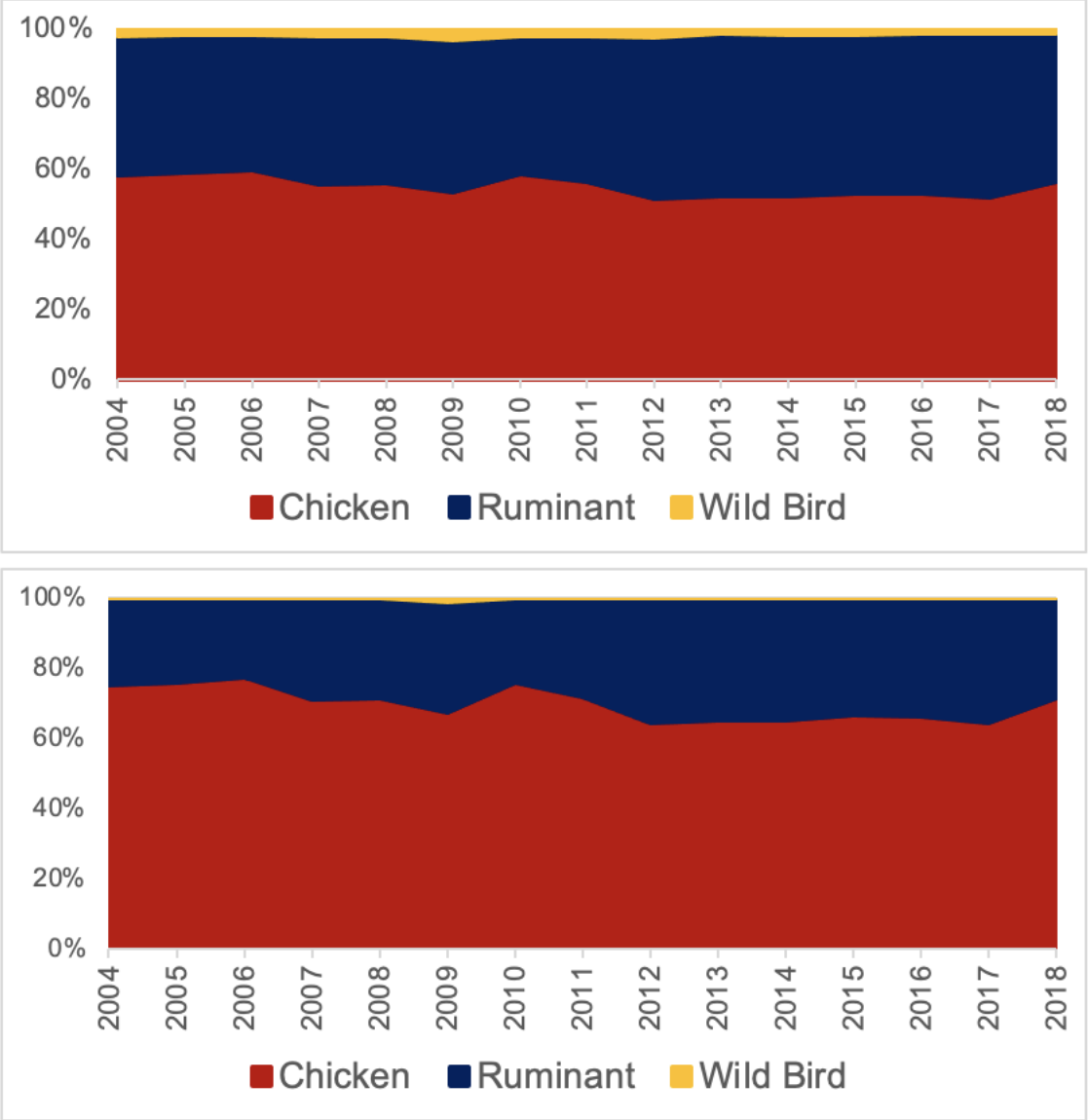



Figure 21. Attribution of human isolates of *Campylobacter jejuni* in Oxfordshire October 2003-September 2018 to chicken, ruminant and wild bird sources by year using the Structure algorithm and MLST data from the source animal reference set (Table 31). The upper panel shows raw results and the lower results adjusted for bias identified in self-attribution. Isolates analysed each year varied between 445 and 813. Study years run from October to the following September and are identified by the later year.

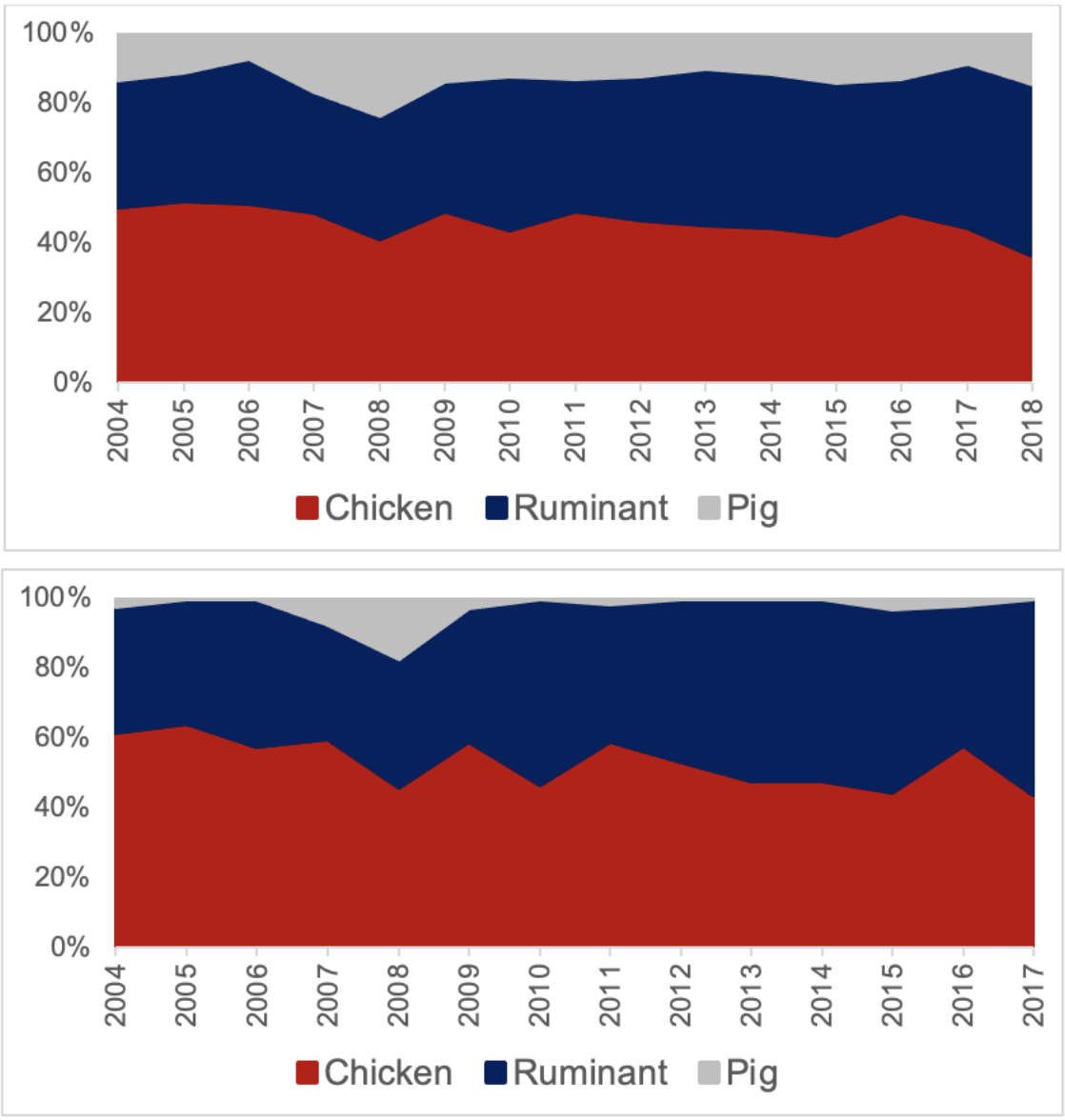



Figure 22. Attribution of human isolates of *Campylobacter coli* in Oxfordshire October 2003-September 2018 to chicken, ruminant and pig sources by year, using the Structure algorithm and MLST data from the source animal reference set (Table 31). The upper panel shows raw results and the lower results adjusted for bias identified in self-attribution. Isolates analysed each year varied between 31 and 107. Study years run from October to the following September and are identified by the later year.

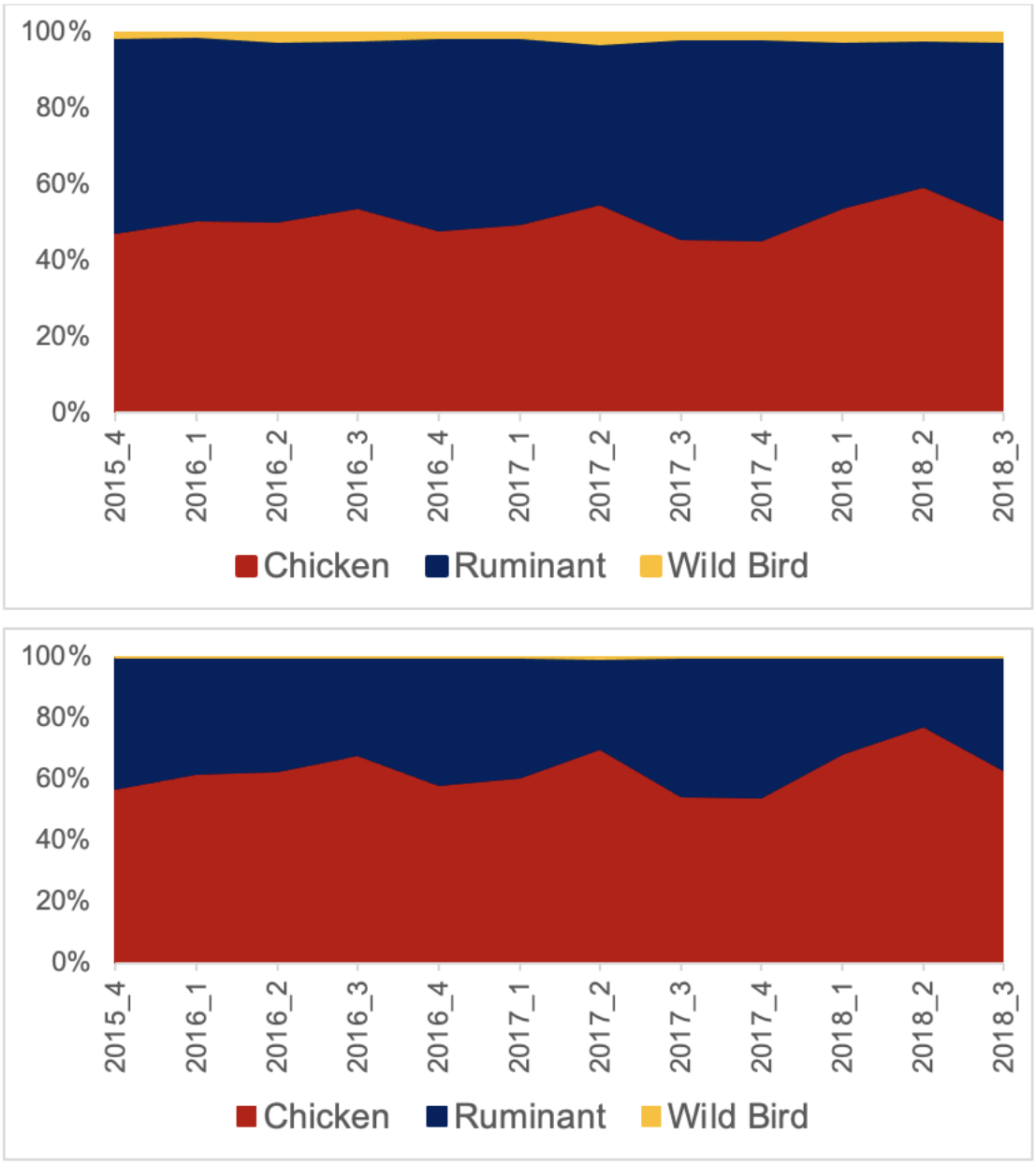



Figure 23. Attribution of human isolates of *Campylobacter jejuni* from North East England and Oxfordshire, by quarter from October 2015 to September 2018 among chicken, ruminant and wild bird sources using the Structure algorithm and MLST data from the source animal reference set (Table 31). The upper panel shows raw results and the lower results adjusted for bias identified in self-attribution. Isolates analysed for each quarter varied between 203 and 449.

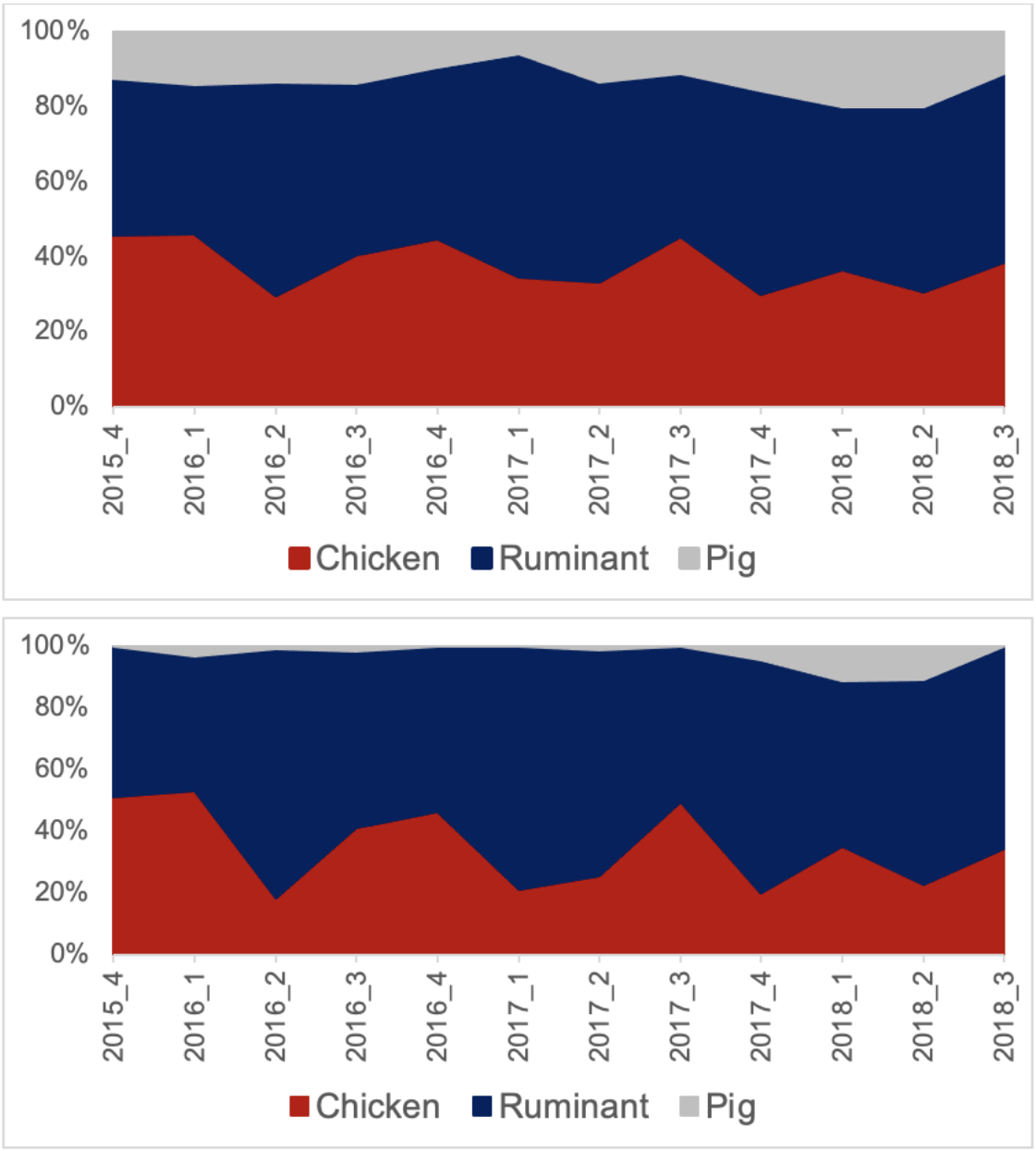
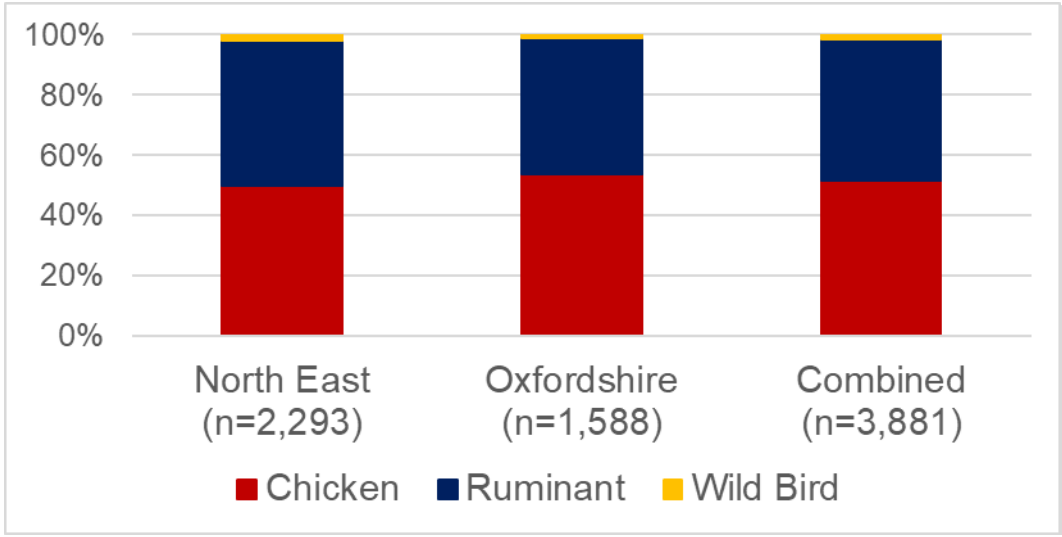
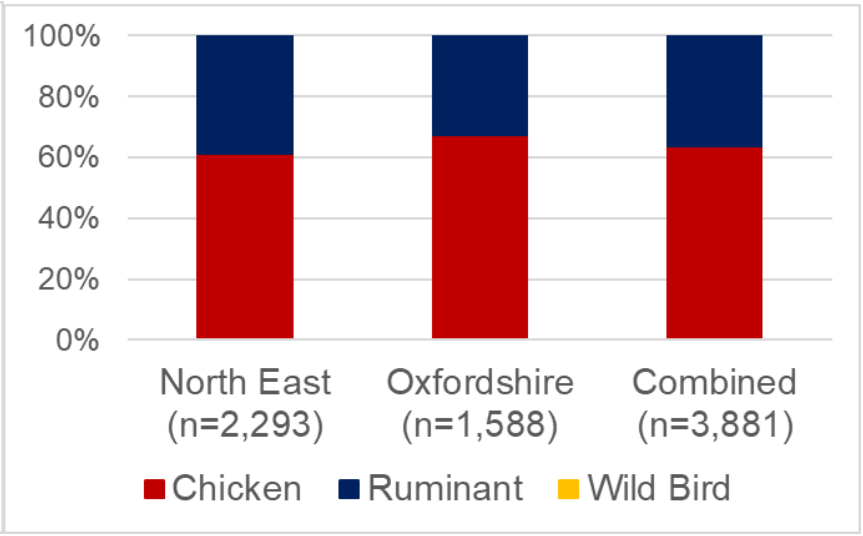
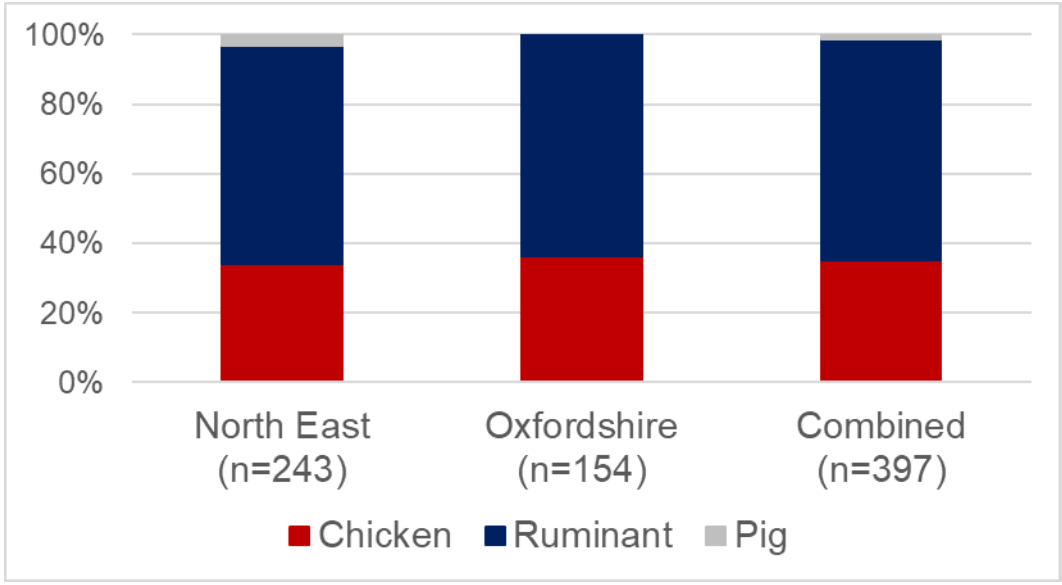



Figure 24. Attribution of human isolates of *Campylobacter coli* from North East England and Oxfordshire, by quarter from October 2015 to September 2018 among chicken, ruminant and pig sources using the Structure algorithm and MLST data from the source animal reference set Table 31). The upper panel shows raw results and the lower results adjusted for bias identified in self-attribution. Isolates analysed for each quarter varied between 17 and 47.


100%
80%
60%
40%
20%
0%
North East (n=2,293)
Oxfordshire (n=1,588)
Combined (n=3,881)
Chicken
Ruminant
Wild Bird


100%
80%
60%
40%
20%
0%
North East (n=2,293)
Oxfordshire (n=1,588)
Combined (n=3,881)
Chicken
Ruminant
Wild Bird


100%
80%
60%
40%
20%
0%
North East (n=243)
Oxfordshire (n=154)
Combined (n=397)
Chicken
Ruminant
Pig

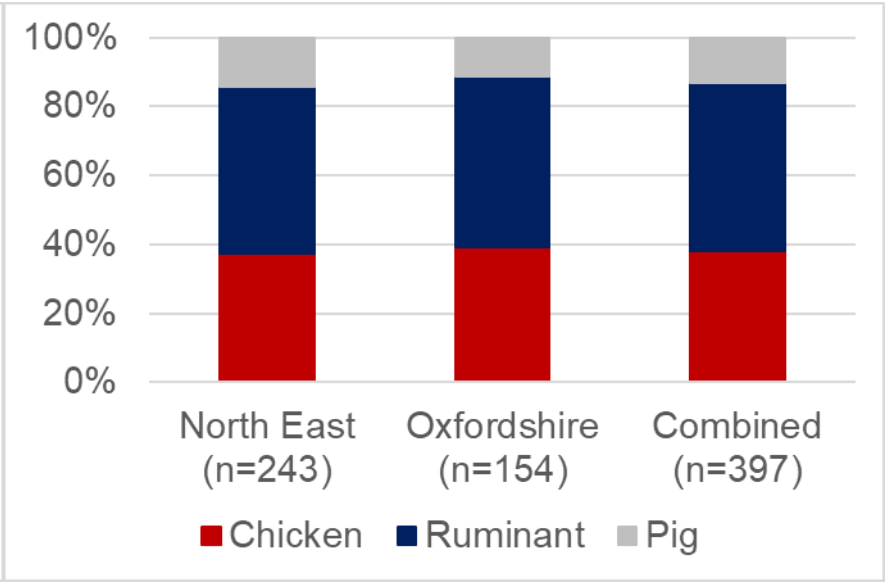
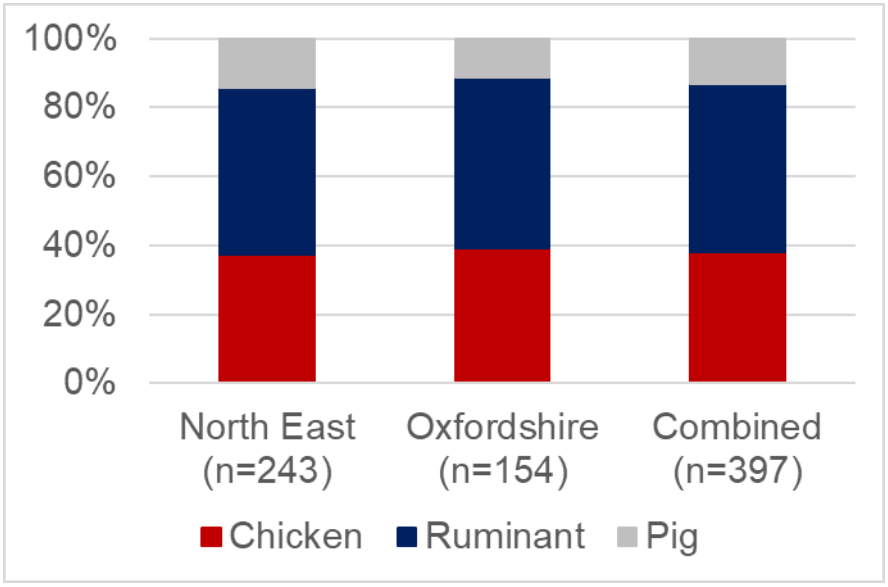
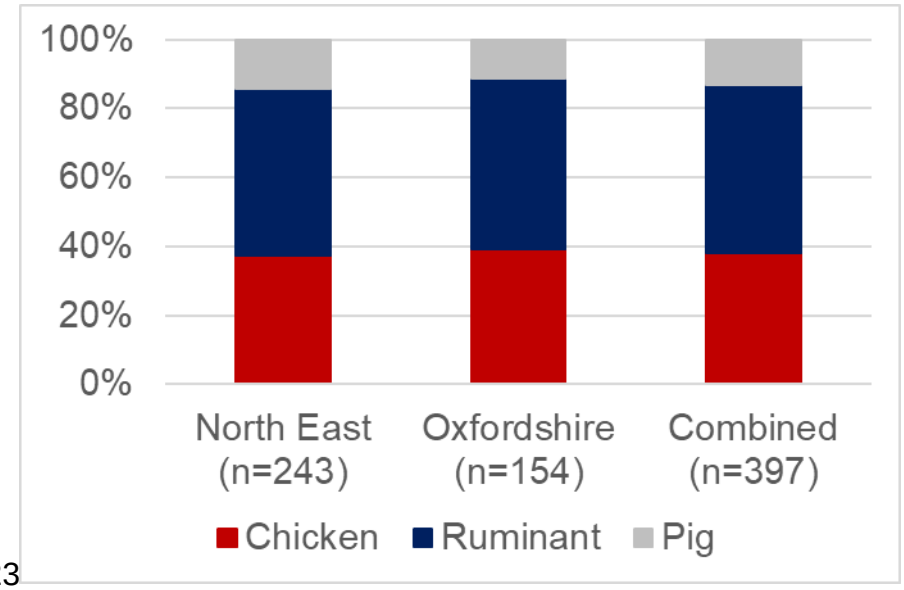

100%
80%
60%
40%
20%
0%
North East (n=243)
Oxfordshire (n=154)
Combined (n=397)
Chicken
Ruminant
Pig

Figure 25. Raw (left hand panels) and adjusted (right hand panels) percentage attribution of isolates from the North East and Oxfordshire sentinel sites October 2015 to September 2016 for *Campylobacter jejuni* (upper panels) and *Campylobacter coli* (lower panels).

For *C. jejuni* the annual attribution across sources is relatively stable across the 15 years reported. Adjustment for bias estimated from self attribution increases the proportion of infection attributed to chicken, reduces the proportion attributed to ruminants and removes most disease attributed to wild bird sources (Figure 21). The findings in the current 2015-2018 sentinel dataset are similar(Figure 23, Figure 25), with an unadjusted estimate of 51.5% attributable to poultry and adjusted of 63.7%. Across the two sites there is a slightly larger proportion of infection attributed to chicken in Oxfordshire (adjusted 67.2% vs 61.3%) with ruminant higher in North East England adjusted 38.7% vs 32.8%) (Figure 23 Figure 25 and Appendix Table 54).



*Campylobacter coli* attribution favours ruminants as the main source, followed by chicken and then pig. Adjustment amplifies these proportions further (Figure 22, Figure 24,

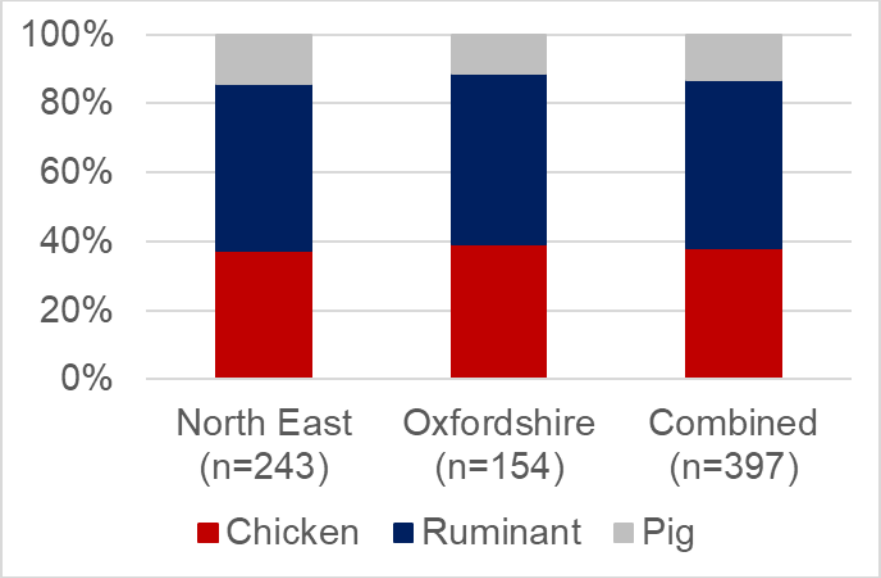



Figure 25). There is stronger evidence for infection originating from pig sources in the North East than in Oxfordshire (

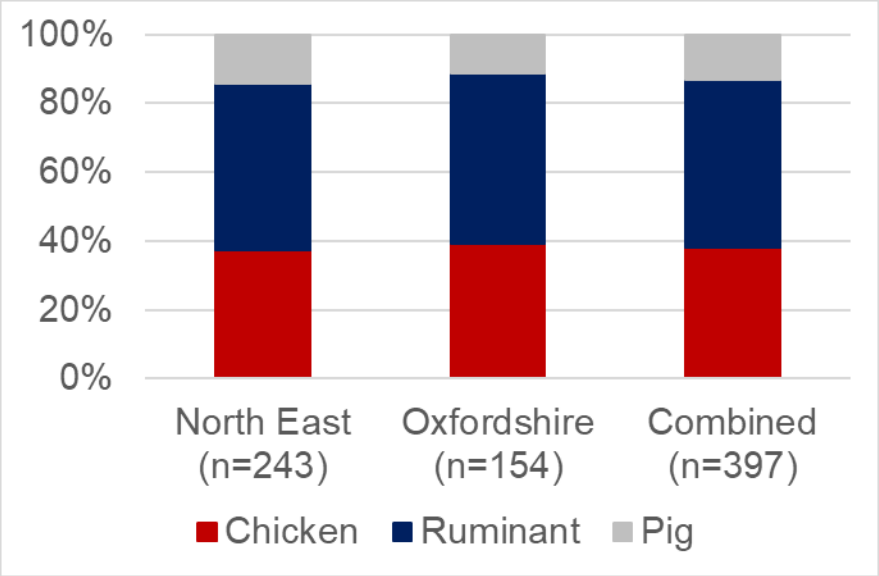



Figure 25). The quarterly attribution results for *C. coli* are prone to substantial stochastic variability with as few as 17 case being attributed per quarter (Figure 24).

## 6.7 Attribution based on sentinel types

The populations of *C. jejuni* in different animal species show partial separation. Using 7 gene multi-locus sequence typing (MLST) many sequence types are shared across hosts. However some MLST types are uniquely, or almost uniquely, found in humans and only one genus (e.g.

chicken or cattle) or sub-order (e.g. ruminant) among non-human animals. The idea of sentinel attribution is to see the proportion of human disease that can be explained by, for example, the human and “chicken only” sequence types and similarly for other host restricted sequence types. To establish the feasibility of an approach using sentinel types all sequence types comprising more than 10 isolates in the *C. jejuni* attribution dataset (Table 31) were assessed to allow identification of these host restricted sentinel types. A range of cut off points were used such as sequence type 90% restricted to a single host, 95% restricted, and approaches considering clonal complex. Estimation took account of the different denominators in the dataset and are based on meeting this threshold adjusted for these different total numbers. The 90% cut off was chosen as optimising sensitivity and specificity of this approach. Specificity ranged from 95% for ruminants (5% of isolates in a type identified as ruminant were actually from chicken or wild bird) to 99% for chicken. There were no sheep specific sentinel types. The sentinel types covered 39% of chicken isolates, 21% of cattle and 36% of ruminant. A 95% cut off lost sensitivity for ruminants (reducing coverage from 36% to 24%) without substantially improving specificity. The sequence types meeting these sentinel criteria are given in the Appendix 8 (Table 56).

Identification of the proportion of human disease caused by these sentinel types for each species, followed by extrapolation based on the proportion of each source species comprising sentinel types (for example 36% of ruminant isolates) then allows estimation of the contribution of each source to human disease. Arithmetically this involves identifying the proportion of human disease that arises from these sentinel types and multiplying this by, for example for chicken 2.56 obtained as 1÷0.39 since 39% of chicken isolates are one of the sentinel types and 61% are not. This multiplier allows extrapolation from sentinel types to the proportion due to all isolates transmitting to humans from the chicken source. Similarly for other sources such as ruminants. These estimates all involve the assumption that these host restricted sentinel types of *Campylobacter jejuni* isolates, transmit to humans at the same rate as types found across multiple host species. The results for the proportion of human infection due to sentinel types are shown in Figure 26 and estimates of the total proportion from each source by extrapolation in Figure 27.

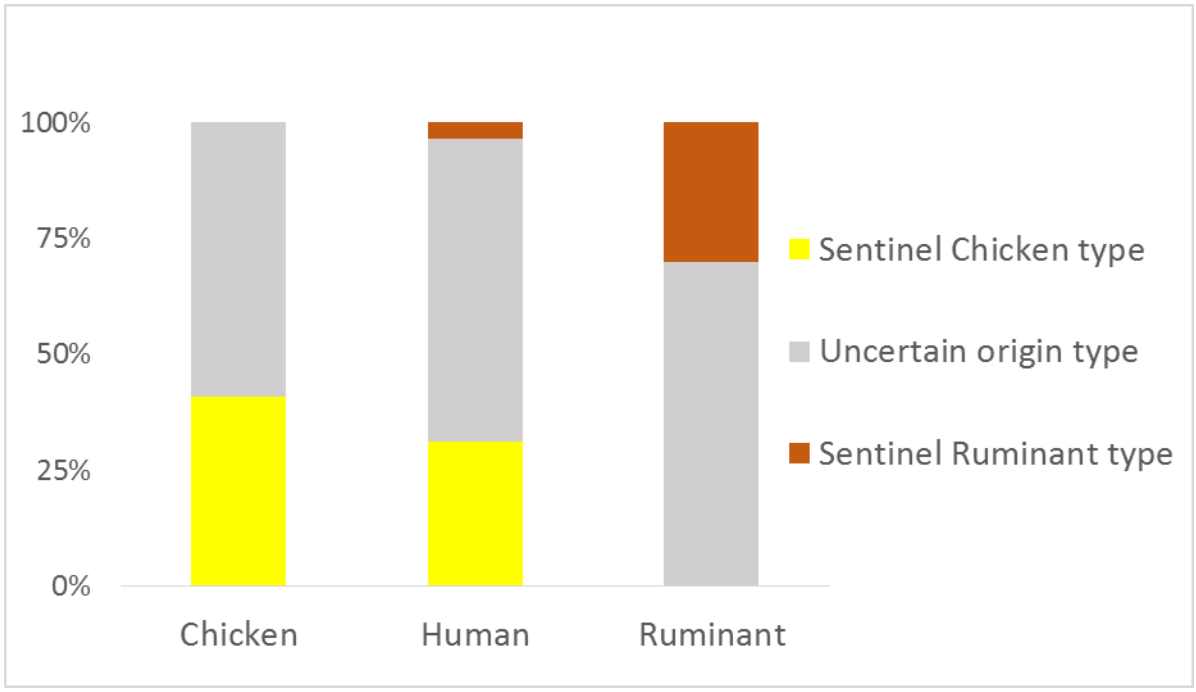



Figure 26. Percentage of *C. jejuni* isolates in humans, chicken, and ruminants due to chicken and ruminant sentinel types, i.e. types identified (almost) exclusively from these sources and humans.

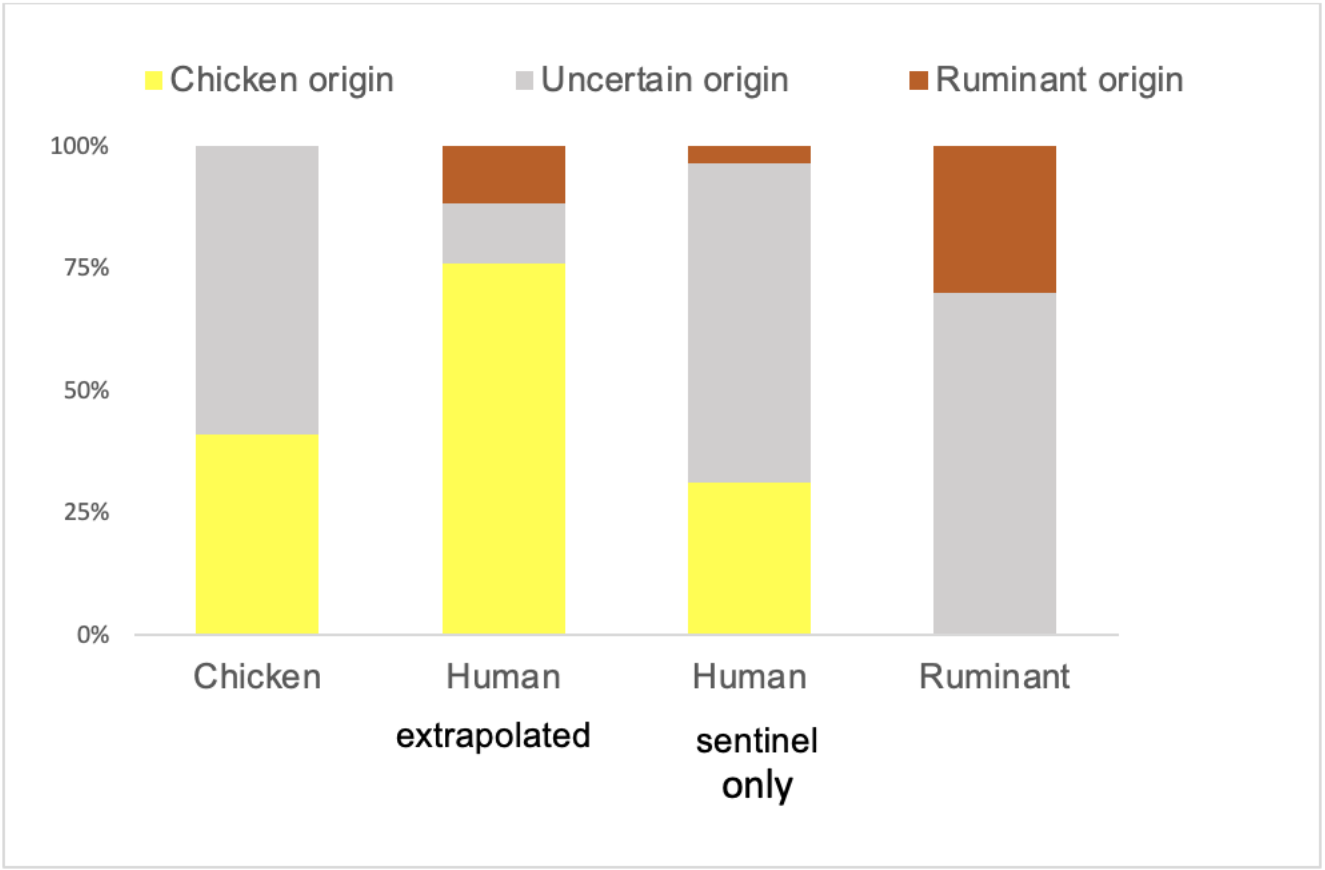


Figure 27. Percentage of *C. jejuni* isolates in humans, chicken, and ruminants due to chicken and ruminant sentinel types, identified (almost) exclusively from these sources and humans and extrapolation to predict the overall proportion of human disease from chicken and ruminant origins. The proportion of human disease attributed to chicken by this method is 75.9% somewhat higher than estimates by adjusted attribution using the Structure algorithm (63.7%). The ruminant proportion (11.9%) is lower than estimates from Structure (36.3%). Figure 28 considers cattle specific subtypes within the ruminant category. Most human infections with ruminant sentinel types map onto types shared by cattle and sheep rather than to those exclusive to cattle.

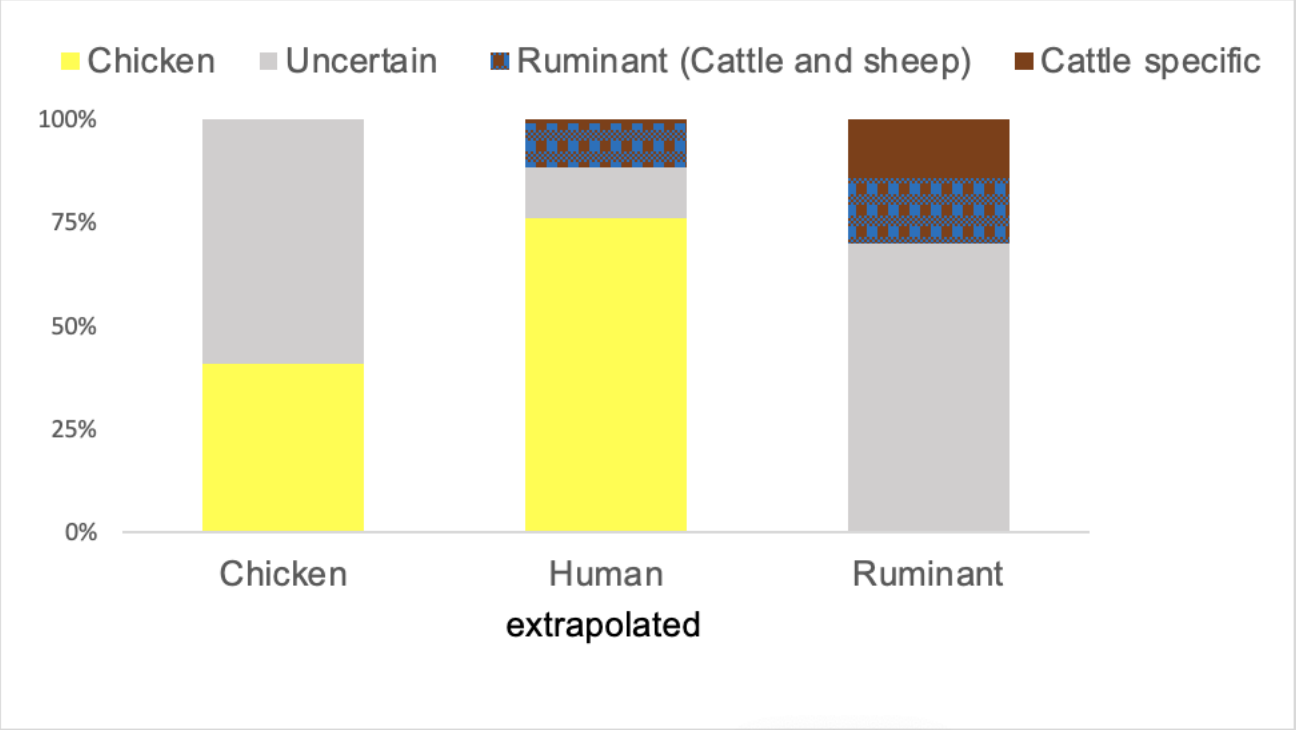



Figure 28. Proportions of chicken and ruminant origin *C. jejuni* due to sentinel types dividing ruminant into cattle restricted and those shared across cattle and sheep. Extrapolation estimates the proportion from sources in humans, again splitting cattle from cattle and sheep types.

## 6.8 Discussion and conclusion

Systematic review of populations genetic approaches to attribution showed between 44% and 83% is human disease as linked to poultry sources. Validation studies showed the potential importance of bias in population genetic attribution of *Campylobacter* to date and the absence of approaches to adjust for it or perform sensitivity analysis. Developing approaches, based on estimates of self-attribution accuracy, to adjust for this bias increased the proportion attributed of *C. jejuni* attributed to chicken sources and *C. coli* attributed to ruminant. A sentinel approach to attribution of human *C. jejuni* infection estimated an even higher proportion attributable to chicken (76%) than these bias adjusted estimates, and identified that cattle restricted sequence types are less common in humans than sequence types shared across sheep and cattle. The patterns of attribution were relatively stable over time with no evidence for substantial or persistent temporal trends in the proportions due to different sources.

Overall this work supports consistent contribution of approximately 70% from chicken sources to human *C. jejuni* infection. The route of transmission cannot be evaluated from these data. The

approaches to adjustment proposed here are likely to continue to be important as attribution moves to whole genome data in the absence of an accurate and unbiased method to date. More highly resolved sentinel approaches, such as from using whole genome data, may also be informative as large whole genome sequenced datasets emerge.

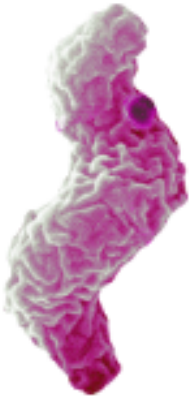
# 7. Data visualisation

## 7.1 A *Campylobacter* Storyboard

A Visual Analytics *Campylobacter* ‘Storyboard’ has been developed that enables users to gain insight into, and understanding of, one of the PubMLST datasets from the project. The PubMLST datasets are large, complex, heterogeneous and multi-dimensional, which means that they are very difficult to assimilate and understand using conventional approaches to querying and examining databases. Storytelling provides an effective means of communicating information that enables users to gain insight and understanding, i.e. to gain Situation Awareness. Storytelling provides a high-level contextual view which is a narrative of what is happening. The Storyboard we have developed uses Visual Analytic techniques to present the story in the data in a manner that enables the user interactively to analyse the infection events and use deductive reasoning to gain understanding, or Situation Awareness, from the dataset. The research and development challenge is to work out how to transform the non-visual data into a visual form that is natural, intuitive and easily accessible for users of all types and knowledge levels. This requires in depth understanding of the nature of the data and the needs of the wide range of potential users. The web-based *Campylobacter* ‘Storyboard’ enables users of all levels to explore and gain understanding of the information contained within a *Campylobacter* isolate database in an easy and yet very detailed manner, and without any training or user manual. The users can thus be the general public, as well as policy makers, decision makers, epidemiologists, researchers, genomics specialists and others.

**The Maiden Lab**
Department of Zoology, University of Oxford





Human campylobacteriosis is very common worldwide and its reduction is an international public health priority. It remains the most common form of bacterial gastroenteritis in the UK, with an annual cost estimated to be as high as £500 million.

The disease comprises a range of symptoms mostly being mild to severe gastroenteritis, but rare forms of paralysis (neuropathies) are among the most serious consequences of this infection.

Campylobacteriosis is caused by infection with bacteria belonging to the genus Campylobacter. Although there are 35 different species of Campylobacter known only one, called Campylobacter jejuni, causes most human disease. Another, called Campylobacter coli is responsible for about 10% of human cases. Both of these bacteria are zoonotic pathogens, which means that they are naturally found in domestic and wild animals from which they can infect and cause disease in humans. Human infection is often the result of eating contaminated food, particularly poorly cooked chicken, but there are other ways of getting the disease, for example from contaminated water or milk.

Picture source: https://commons.wikimedia.org/wiki/File:Campylobacter_jejuni.png

The following shows the summary of the infections in female and male that happened during October 2015 - April 2018.


The following shows the temporal patterns of the infections in male and female.

## Age


## Summary

**Background information**
Campylobacter jejuni and Campylobacter coli are the most frequent causes of acute bacterial gastroenteritis in humans in the UK and represent an unrelenting worldwide public health problem. C. jejuni accounts for over 90% of cases, with the majority of the remainder caused by C. coli.

The annual incidence of diagnosed human infections in 2008 was 92 per 100,000 individuals in England and Wales.

**Campylobacter**
In the UK, the incidence of Campylobacter infection varies with age, being highest among those under five years of age. Males are more frequently affected, with a male-to-female ratio of 1.2 to 1.

Many wild and farmed avian and mammalian species carry campylobacters as natural and apparently harmless members of their commensal gastrointestinal microbiota. Human infection results from either direct contact with contaminated faeces or indirect transmission via contaminated food or drink.

Outbreak investigation, case control, and natural experiments have implicated contaminated poultry meat as a major cause of clinical infection. Our visualisations allow exploration of the characteristics of reported cases associated with bacterial isolates.

## Treatment

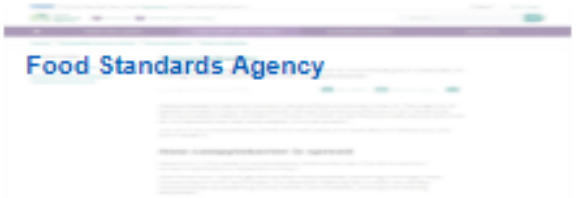

## Gender


## Publications


## Precautions

**How to avoid Campylobacter poisoning?**
The Food Standards Agency recommends:

1. Cover and chill raw chicken;
2. Don't wash raw chicken;
3. Wash hands and used utensils;
4. Cook chicken thoroughly.




## Campylobacter Species


## Data

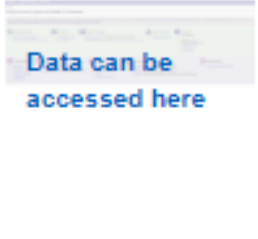


Figure 29: Screenshot of the interactive *Campylobacter* Storyboard waiting to be released publicly.

## 7.2 A *Campylobacter* Dashboard

A *Campylobacter* Dashboard was developed to apply visual analytics to analyse data on the contamination of whole fresh chickens from the Food Standard Agency retail chicken survey (FS102121) using the available data for the period 8th July 2015 – 30th March 2016. There were 2998 chickens which tested positive for *Campylobacter* and among these 342 chickens had more than 1,000 colony forming units (cfu) - and were thus considered to be highly contaminated.
The Dashboard shows the geographic distribution of the eleven retailers involved in the study. The two Stacked Column charts on the top right-hand side show the distribution of cfu counts on the neck skin and the outer packaging swab, colour coded by the detected *Campylobacter* species. Among neck skin samples 11.4% of the contaminated chickens had more than 1,000 cfu. There were five with more than 1,000 cfu in their outer packaging: these were sampled at different times, different locations, and from different retailers as well as coming from different slaughterhouse premises (different processor approved numbers). The retailer and suppliers Stacked Column chart shows that most retailers are supplied by a major supplier with additional secondary suppliers used to lesser extents. The temporal Stacked Column chart shows the temporal pattern of the chickens - colour coded by the types of chicken, i.e. standard, free range or organic. The dashboard provides an effective means of analysing the temporal and geospatial correlation, patterns and trends in the observed cfu data.

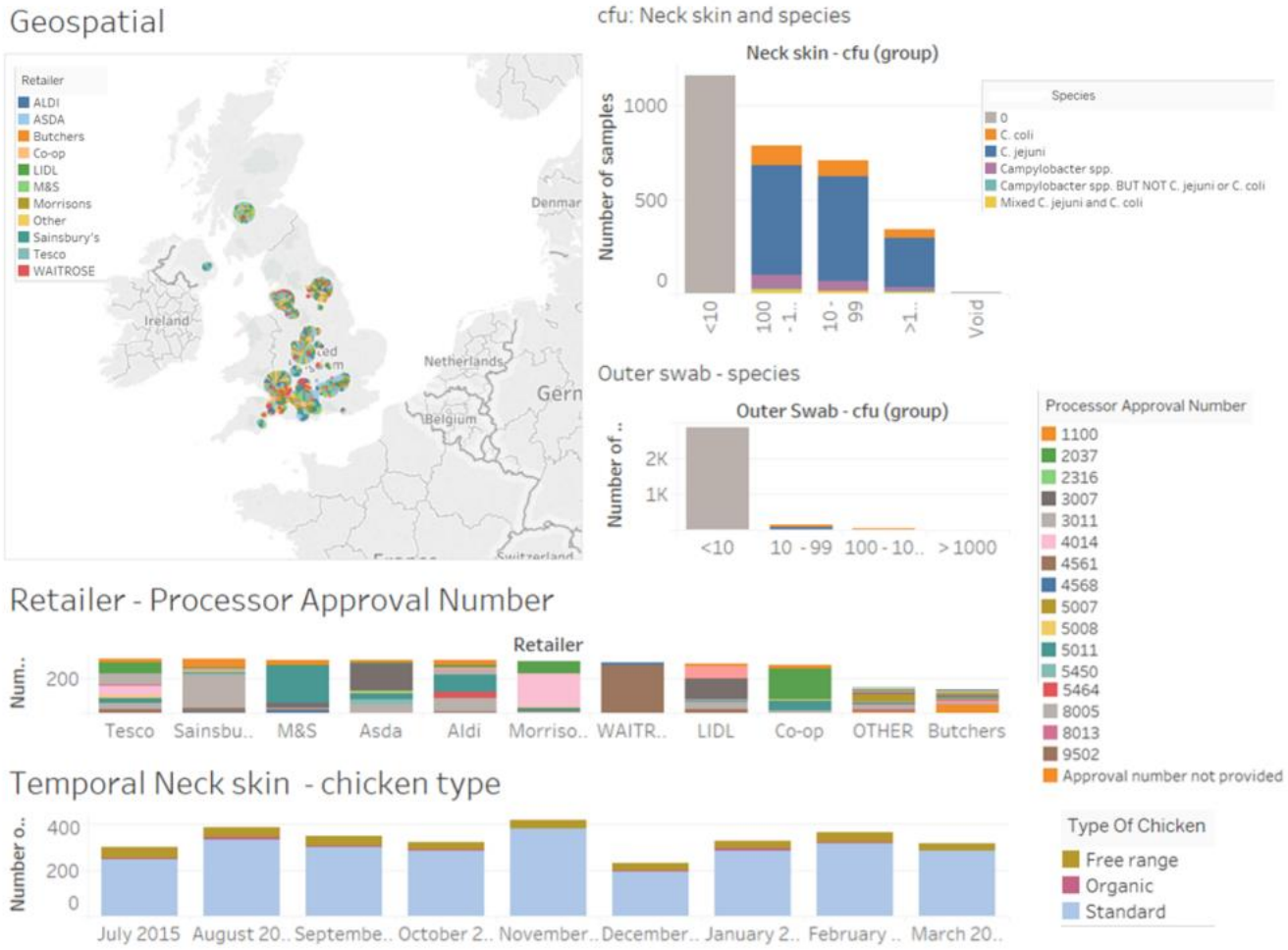



Figure 30. Screenshot of an interactive example of outputs from investigations into *Campylobacter* in fresh whole retail chickens.

# 8. Antimicrobial resistance

## 8.1 Method

Antimicrobial resistance was estimated using a previously validated and published in-house bioinformatics pipeline in PHE to predict antimicrobial resistance profiles from genome sequence data to macrolides, fluoroquinolones, tetracycline and aminoglycosides (62). This involves the detection of genes by mapping of short-reads to reference genomes for each clonal complex with the full method and validation study currently submitted for publication. This was applied to the clinical and food animal isolates of *C. jejuni* and *C. coli* from this study: sequence reads from 4,362 *C. jejuni* and *C. coli* isolated from recent cases of human infections at the two Sentinel Sites, 2,007 recent isolates from food sources (ruminant liver, duck, pig and chicken) and 1,245 archived isolates from humans. The study did not include phenotypic testing so that this validated genome based resistance assay is our primary measure of resistance.

Additionally, for macrolide, tetracycline and fluoroquinolone antibiotics analysis of *de novo* genome assemblies in the PubMLST campylobacter database using the Bacterial Isolate Genome Sequence Database (BIGSdb) software was also applied to detect and record either the presence or absence of acquired antimicrobial resistance genes, or resistance conferring mutations in core genome loci and results were compared to the PHE result. Although not formally validated in a public health reference laboratory the results were highly similar to those from the PHE validated assessment of resistance with between 0 and 0.4% discordance between results of the two methods across these three antimicrobial classes. This supports the use of this approach with BIGSdb on genome data present on PubMLST. The particular advantages of this are that it is publicly available and that it allows joint analysis of these antimicrobial resistance determinants in the context of the population genetic structure of *Campylobacter* estimated from whole genome multilocus sequence data.

The PHE method results are reported in this section. The BIGSdb results for resistance and comparison are reported in Appendix 9. Results for each isolate from the current study are also recorded on PubMLST and details on how to access these also given in Appendix 9.

## 8.2 Results

Amongst *C. jejuni*, fluoroquinolone and tetracycline resistance in clinical isolates, assayed from their genomes, was highest in the recent years covered by this study (2015-2018) (45.1% and 42.8% respectively) and lower in those isolates from 1997-98 (5.2% and 21.0% respectively).

Comparatively, resistance to macrolides amongst *C. jejuni* has remained low reaching its highest level of 0.4% in recent years (Figure 31and tabulated in Appendix 9, Table 58).

Amongst the *C. jejuni* isolated from food sources, resistance to fluoroquinolone was high in chicken (retail 52.4%, abattoir 42.5%) and lower in ox/calf liver (13.1%), duck (9.1%) lamb liver (8.0%) and pig (0%). Tetracycline resistance in *C. jejuni* was high in pig (70.0%), chicken (retail 60.6%, abattoir 55.9%) and duck (59.1%) isolates, and lower in ox/calf liver (14.1%) and lamb liver (6.6%). Macrolide resistance was not detected in any of the isolates of *C. jejuni* from food sources (Figure 32, and tabulated in Appendix 9).

For clinical isolates of *C. coli,* antimicrobial resistance to fluoroquinolones and tetracyclines rose from low levels in 1997/8 although estimated from a small sample (6% for each, 95% confidence interval 0-27%) to higher levels since (tetracycline 35%-43%, and fluoroquinolones 28%-38% across 2003-2018 samples). The larger sample in the 2015-2018 collection makes these estimates more precise (tetracycline 38% 95% confidence interval 34%-42%, and fluoroquinolones 37%, 95% confidence interval 33%-41%). Although lower than for tetracyclines and fluoroquinolones, resistance levels to aminoglycosides and macrolides were higher than for *C. jejuni*. (Figure 33, and tabulated in Appendix 9).

Amongst the *C. coli* isolated from food sources, resistance to fluoroquinolone was high in isolates from duck (54.5%) and chicken (retail 48.1%, abattoir 50.7%) and lower in ox/calf liver (20.0%), pig (16.9%) and lamb liver (14.3%) isolates. Tetracycline resistance in *C. coli* was highest in pig (85.5%) and duck (81.8%), lower in chicken (retail 60.2% and abattoir 63.8%) and lowest in lamb liver (14.3% and ox/calf liver (10%) isolates. Macrolide and aminoglycoside resistance was highest in pig isolates (33.7% and 59.0% respectively) (Figure 34, and tabulated in Appendix 9).

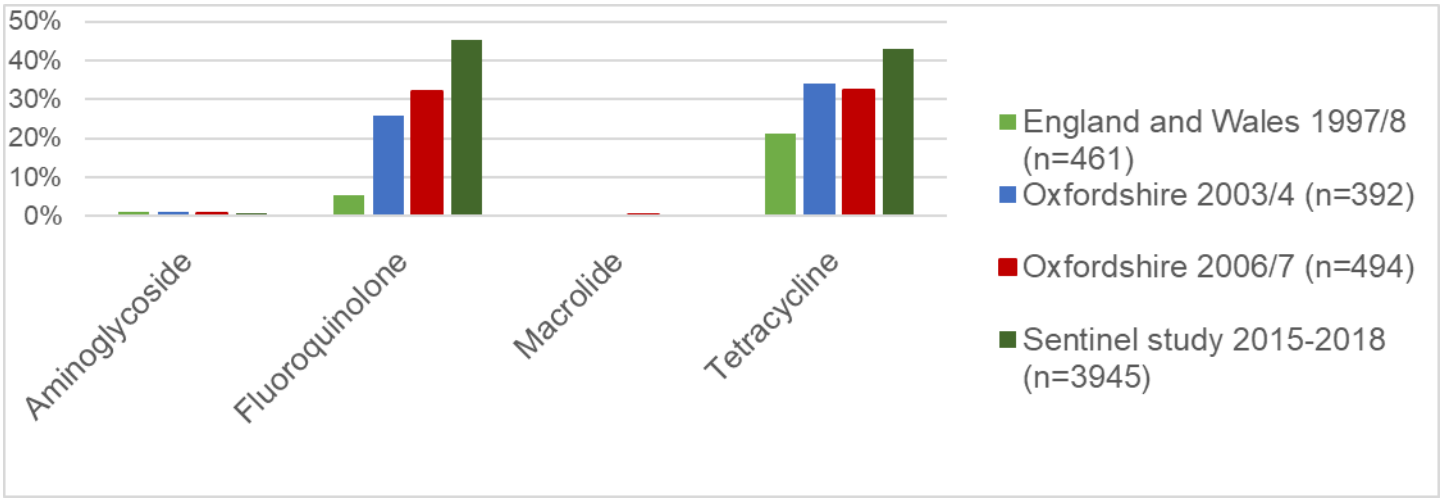



Figure 31: Percentage genetically estimated antimicrobial resistance in isolates of *Campylobacter jejuni* from human cases in datasets spanning 1997 to 2018 in England and Wales.

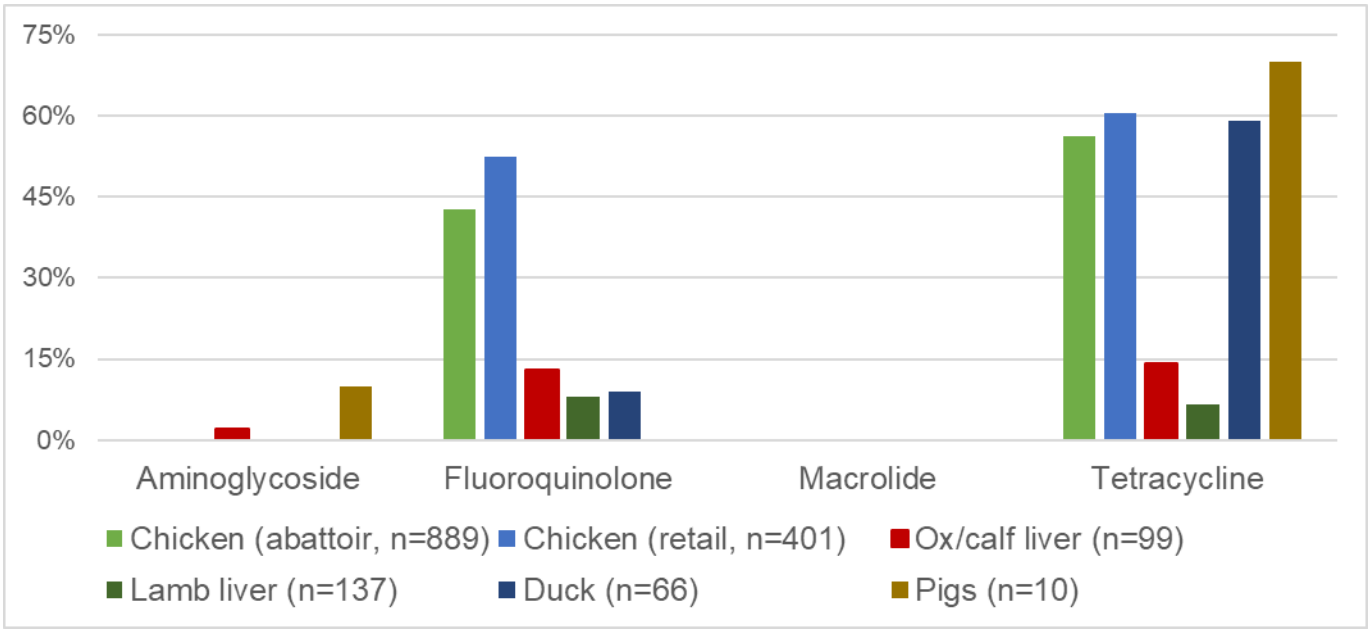



Figure 32: Percentage genetically estimated antimicrobial resistance among isolates of *Campylobacter jejuni* from food animals.

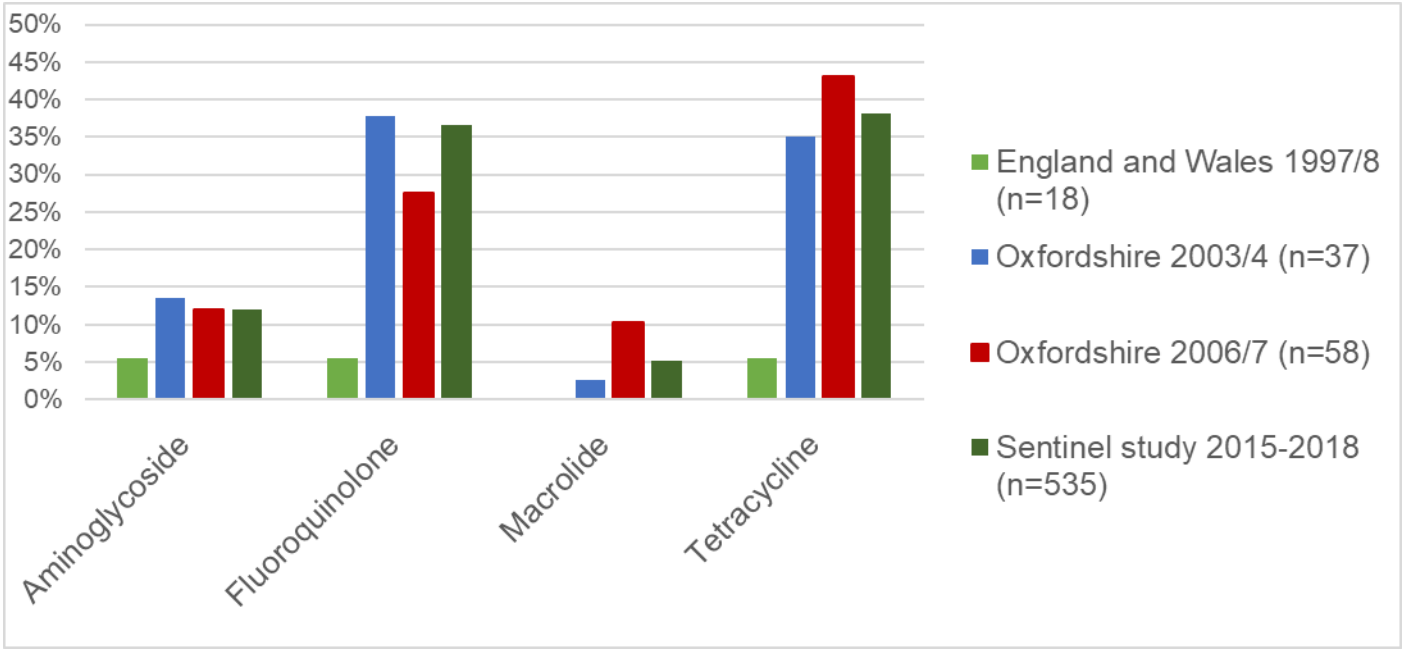



Figure 33: Percentage genetically estimated antimicrobial resistance among isolates of *Campylobacter coli* from human cases in datasets spanning 1997 to 2018 in England and Wales.

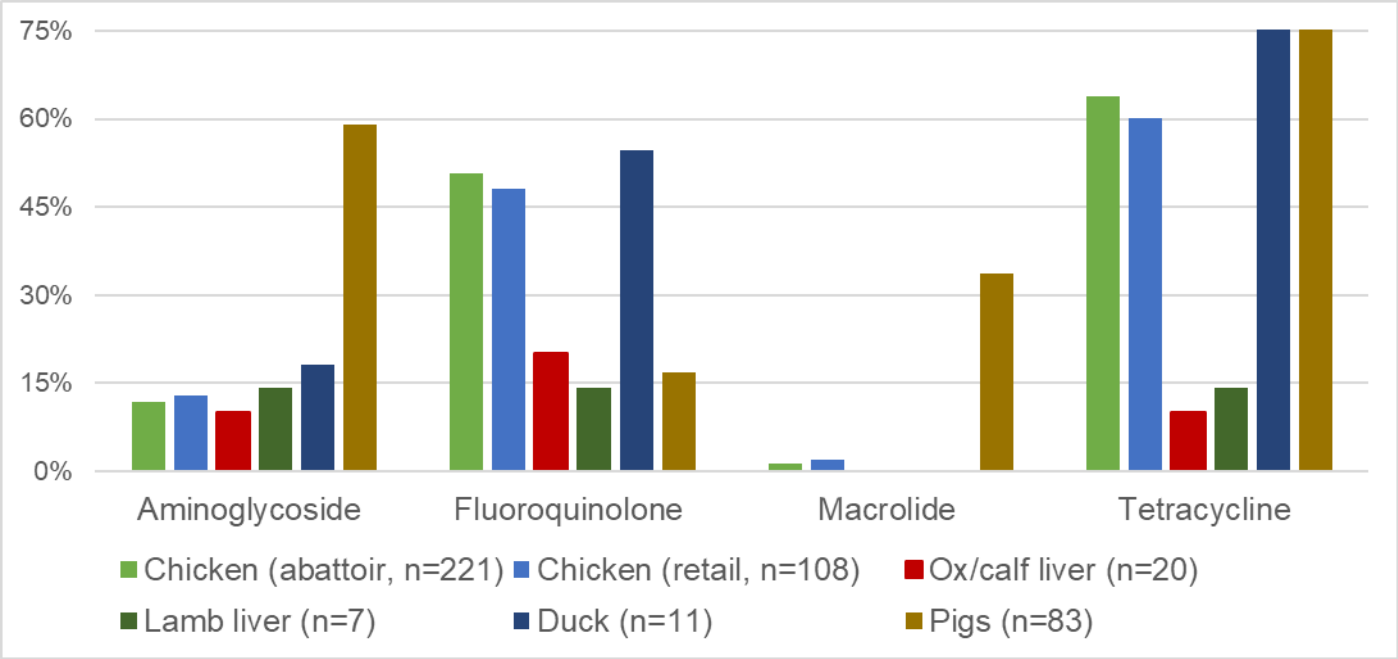



Figure 34: Percentage genetically estimated antimicrobial resistance among isolates of *Campylobacter coli* from food animals.

Graphs for equivalent data using BIGSdb are included in Appendix 9. The BIGSdb analysis also describe the distribution of antimicrobial resistance across clonal complexes among clinical isolates from 2015 to 2018 as presented below. The analysis is restricted to clonal complexes represented by ten or more clinical isolates in the dataset.

Table 36. The association of ciprofloxacin resistance with clonal complexes of clinical *Campylobacter* isolates collected between 1st October 2015 and 30th September 2018. Significant positive association with the resistance for that antibiotic ($p < 0.01$ on Chi square test or Fisher's exact test).

| **Clonal complex** | **Resistant** | **Sensitive** | ***P* value** |
|---|---|---|---|
| ST-1034 complex | 6 | 4 | 0.356 |
| ST-206 complex | 77 | 215 | 0.0001 |
| ST-21 complex | 394 | 541 | 0.096 |
| ST-22 complex | 6 | 78 | 0.0001 |
| ST-257 complex | 100 | 135 | 0.534 |
| ST-283 complex | 1 | 34 | 0.0001 |
| ST-353 complex | 323 | 22 | 0.0001 |
| ST-354 complex | 135 | 13 | 0.0001 |
| ST-403 complex | 20 | 36 | 0.182 |
| ST-42 complex | 10 | 59 | 0.0001 |
| ST-443 complex | 27 | 65 | 0.003 |
| ST-45 complex | 13 | 222 | 0.0001 |
| ST-460 complex | 6 | 6 | 0.702 |
| ST-464 complex | 267 | 11 | 0.0001 |
| ST-48 complex | 31 | 272 | 0.0001 |
| ST-49 complex | 14 | 22 | 0.495 |
| ST-508 complex | 0 | 13 | 0.001 |
| ST-52 complex | 27 | 39 | 0.553 |
| ST-574 complex | 16 | 9 | 0.049 |
| ST-607 complex | 26 | 0 | 0.0001 |
| ST-61 complex | 6 | 72 | 0.0001 |
| ST-658 complex | 26 | 67 | 0.001 |
| ST-677 complex | 0 | 14 | 0.0001 |
| ST-828 complex* | 134 | 227 | 0.003 |
| UA *C. jejuni* | 141 | 72 | 0.0001 |
| UA *C. coli** | 11 | 17 | 0.577 |

UA Unassigned – for isolates not part of any identified clonal complex; *indicates *C. coli* isolates.

Table 37.The association of tetracycline resistance with clonal complexes of clinical *Campylobacter* isolates collected between 1st October 2015 and 30th September 2018. Significant positive association with the resistance for that antibiotic ($p < 0.01$ on Chi square test or Fisher's exact test).

| **Clonal complex** | **Resistant** | **Sensitive** | ***P* value** |
|---|---|---|---|
| ST-1034 complex | 7 | 3 | 0.109 |
| ST-206 complex | 185 | 107 | <0.0001 |
| ST-21 complex | 280 | 655 | <0.0001 |
| ST-22 complex | 2 | 82 | <0.0001 |
| ST-257 complex | 133 | 102 | <0.0001 |
| ST-283 complex | 0 | 35 | <0.0001 |
| ST-353 complex | 243 | 102 | <0.0001 |
| ST-354 complex | 133 | 5 | <0.0001 |
| ST-403 complex | 9 | 47 | <0.0001 |
| ST-42 complex | 12 | 57 | <0.0001 |
| ST-443 complex | 36 | 56 | 0.505 |
| ST-45 complex | 13 | 222 | <0.0001 |
| ST-460 complex | 5 | 7 | 1 |
| ST-464 complex | 264 | 10 | <0.0001 |
| ST-48 complex | 51 | 252 | <0.0001 |
| ST-49 complex | 5 | 31 | <0.0001 |
| ST-508 complex | 13 | 13 | 0.439 |
| ST-52 complex | 19 | 47 | 0.023 |
| ST-574 complex | 25 | 0 | <0.0001 |
| ST-607 complex | 24 | 6 | <0.0001 |
| ST-61 complex | 5 | 73 | <0.0001 |
| ST-658 complex | 17 | 76 | <0.0001 |
| ST-677 complex | 0 | 14 | 0.001 |
| ST-828 complex* | 128 | 234 | 0.004 |
| UA *C. jejuni* | 118 | 96 | <0.0001 |
| UA *C. coli** | 11 | 17 | 0.728 |

Isolates from five clonal complexes (cc) ST-353cc, ST-354cc, ST-464cc, ST-574cc and ST-607cc, together with isolates with sequence types (ST) unassigned to a clonal complex were strongly associated with fluoroquinolone resistance, as determined by Pearson's chi-squared test, or where numbers were fewer than five, by Fisher's exact test. Isolates from seven clonal complexes ST-206cc, ST-257cc, ST-353cc, ST-354cc, ST-464cc, ST-574cc and ST-607cc, together with *C. jejuni* isolates with sequence types (ST) unassigned to a clonal complex were strongly associated with tetracycline resistance. The ST-257cc, ST-353cc, ST-354cc, ST-464cc and ST-607 clonal complexes are generally chicken-associated, whilst ST-206 complex is more commonly associated with ruminants.

## 8.3 Discussion

Overall, these results describe the dramatic rise in fluoroquinolone resistance in isolates from human *C. jejuni* infections since 1997 in keeping with past publications (63), and substantial and also rising tetracycline resistance. Resistance to macrolides and aminoglycosides remains low. Across food animals tested fluoroquinolone resistance was high in chicken and much lower in other sources. Tetracycline resistance was high in chicken, duck and pig isolates and lower in cattle and sheep isolates. This may reflect variation in current and past antimicrobial use across the food industry and contrasts with other countries with lower levels of historical fluoroquinolone use in this setting such as Australia (51).

In *C. coli* the pattern is somewhat different. Aminoglycoside resistance is higher across all time periods among human cases and across animal species, with particularly high levels in isolates from pigs, which also show substantial macrolide resistance. Tetracycline resistance shows similar patterns of resistance across animal species to the pattern for *C. jejuni*. For fluoroquinolones a substantial rise from the earliest human samples (1997/8) is also evident. The pattern across animal species is similar other than a 55% level of resistance in samples from duck.

The large scale and long time period covered by the human component of this study offers a unique view into the emergence of antimicrobial resistance in *Campylobacter* at the genome level. As outlined under further work a doctoral studentship will analyse and model this in more detail.

## 9. Cluster detection

Public Health England, with a remit for the investigation and control of outbreaks of gastroenteritis, assessed the potential utility of genome sequence data to identify clustering suggestive of a common source and therefore a potential support to the detection and investigation of outbreaks. These approaches have been developed and applied to *Salmonella* in the past and used in public health investigations by Public Health England for *Salmonella*.

A parallel analysis was undertaken using core genome multi-locus sequence typing data across data from the current study and data from all other isolates accessible on the PubMLST database which provides a global archive of assembled *Campylobacter* genomes. This assessed the relatedness of isolates from human disease clusters in the current study with international and non-human origin isolates.

### 9.1 Analysis of data from this study to inform PHE protocols for cluster detection

#### PHE single nucleotide polymorphism analysis

Single nucleotide polymorphisms (SNPs) were identified based on ST-complex specific reference mapping and cluster detection was performed across the 13 most prevalent ST-complexes (covering 85.5% of the isolates). Hierarchical single linkage clustering was performed, and 5-SNP single linkage clusters were hypothesised to contain isolates similar to the level expected from a shared source, without any time restrictions. Many isolates were not in a single linkage cluster with any other isolate at the 5-SNP level. These, with their own unique 5-SNP "group" in the dataset, are consistent with a sporadic isolate. As a summary measure to describe overall clustering within a clonal complex or species the N50 was defined as the smallest number of 5-SNP groups that would contain at least 50% of all isolates in that clonal complex or species. The associated P50, the percentage of all observed distinct 5-SNP groups comprising at least 50% of isolates was calculated.

Table 38. Isolates with single nucleotide polymorphism data available across both sentinel sites.

| Sentinel Sites | Isolates | SNP typed | Not SNP typed |
|---|---|---|---|
| North East | 2606 | 2249 – 86.3% | 357 – 13.7% |
| Oxfordshire | 1766 | 1489 – 84.3% | 277 – 15.7% |
| Total | 4372 | 3738 – 85.5% | 634 – 14.5% |

For *C. coli* 382 isolates belonged to 318 distinct 5-SNP groups. The majority of cases were thus apparently sporadic with 91% of the groups containing a single isolate and representing 83% of the *C. coli* cases (Figure 35). The 3,356 *C. jejuni* isolates analysed belonged to 1,868 distinct 5-SNP groups. 81% of these 5 SNP groups included one case accounting for 1,510, apparently sporadic, cases, 46% of the total. Half of *C. jejuni* cases belonged to 15.9% (297/1868) of the 5-SNP clusters (Figure 36). The P50 across the five most common clonal complexes ranged from 18% for ST-21 to 2.8% for ST-353 (

Table 39).

Table 39. Cases, 5-SNP groups (of 1 or more isolates), clusters (2 or more isolates), and the number of clusters containing 50% of all isolates in that clonal complex (N50).

| ST-complex | Cases | Distinct 5-SNP groups | Cluster [2 or more] | N50 cases |
|---|---|---|---|---|
| ST-21 | 956 | 559 | 437 (78%) | 102 (18%) |
| ST-353 | 341 | 145 | 121 (83.4%) | 4 (2.8%) |
| ST-48 | 316 | 175 | 133 (76%) | 30 (17%) |
| ST-206 | 303 | 170 | 131 (77%) | 29 (17%) |
| ST-464 | 290 | 88 | 69 (78.4%) | 3 (3.4%) |

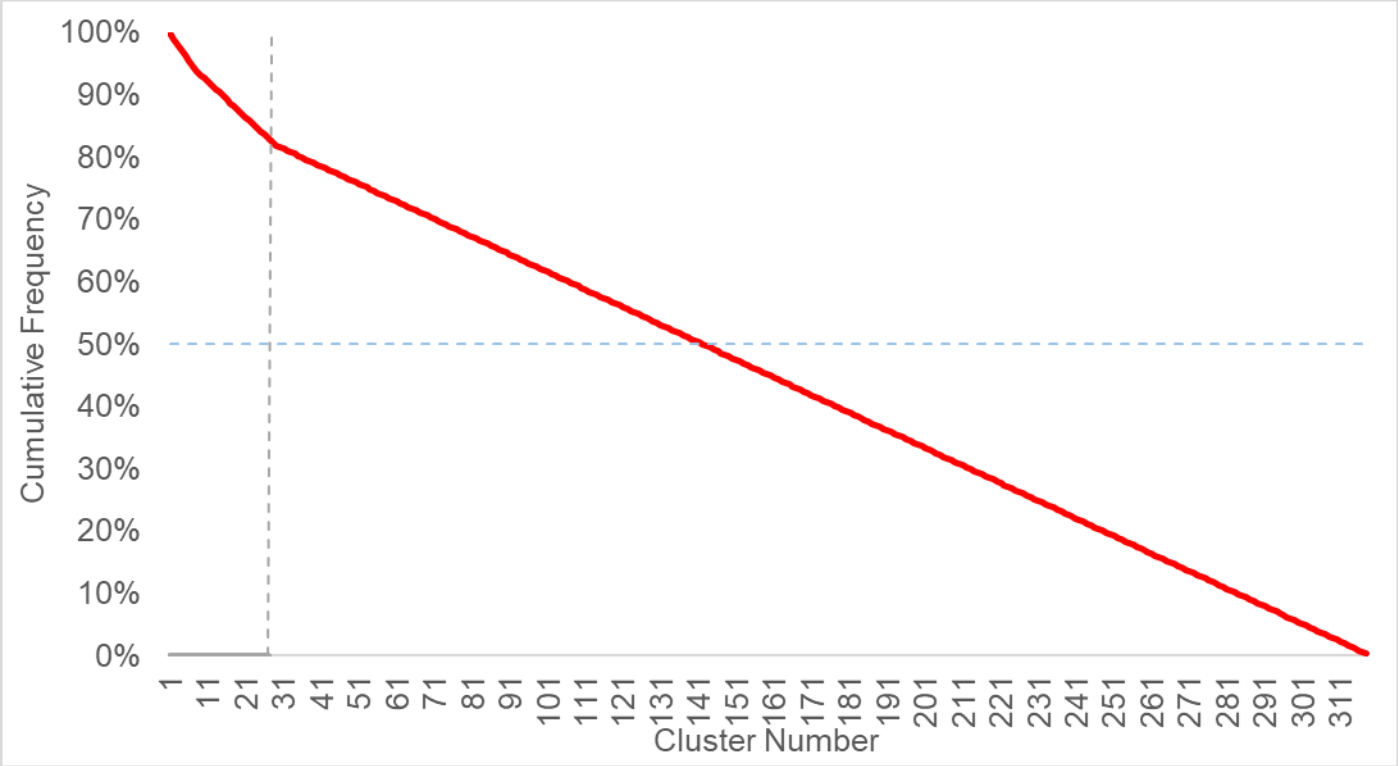



Figure 35. Cumulative frequency of *C. coli* isolates per distinct 5 SNPs groups. Groups are ranked from biggest (n=4) to the smallest (n=1). The horizontal line represents 50% of the cases. Clusters with more than one case are to the left of the vertical line, indicating clustering while those to the right of it are apparently sporadic in this dataset using the 5-SNP threshold.

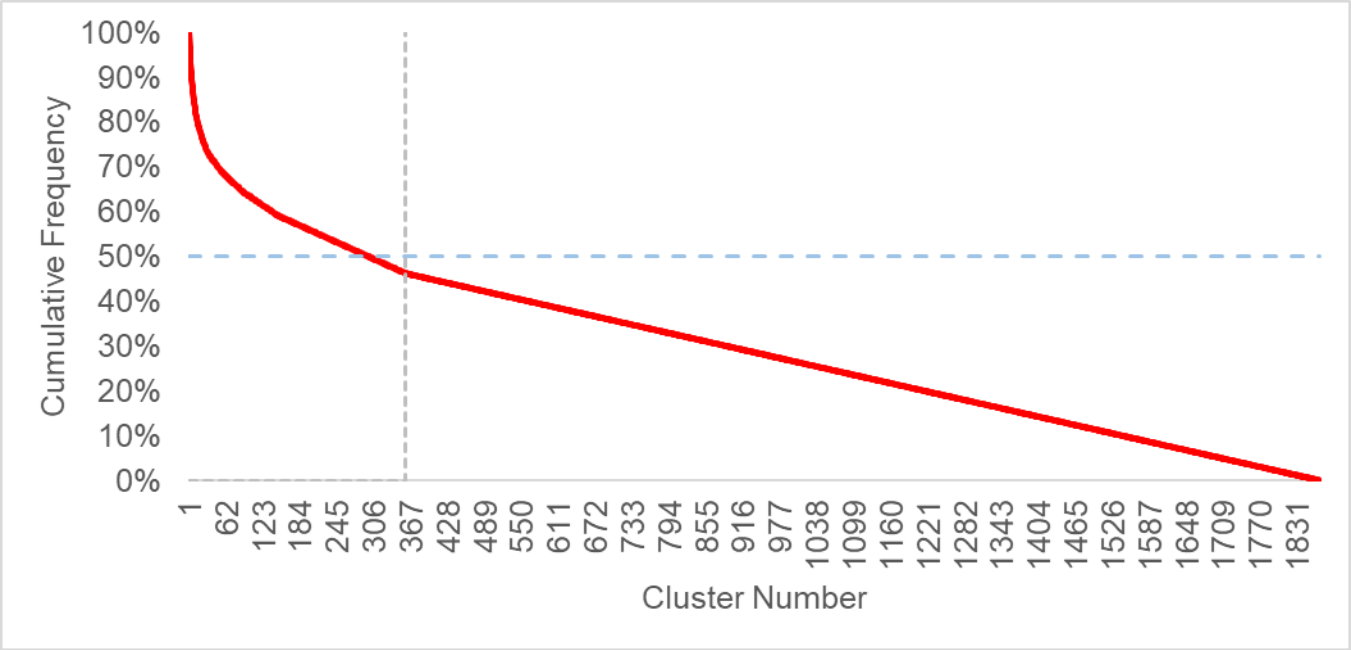



Figure 36. Cumulative frequency of *C. jejuni* isolates per distinct 5 SNPs groups. Groups are ranked from biggest (n=116) to the smallest (n=1). The horizontal line represents 50% of the cases. Clusters with more than one case are to the left of the vertical line, indicating clustering while those to the right of it are apparently sporadic in this dataset using the 5-SNP threshold.

These analyses show marked differences between *C. coli* and *C. jejuni*. For *C. coli* the large majority of isolates (83%) had no closely genetically related isolate identified in the study population, and, where identified, clusters were very small (4 or fewer isolates). For *C. jejuni*, a substantial proportion of isolates (46%) were also apparently sporadic using this threshold. However the remaining 54% were members of clusters ranging from small (2 isolates) to very large (116 isolates). This substantial clustering suggests shared sources or transmission for isolates within clusters that could be the targets for investigation and intervention.

## 9.2 Cluster detection on study data and the PubMLST database using cgMLST

The *C. jejuni* and *C. coli* core genome MLST (cgMLST) scheme (21) has been developed with functionality to detect clusters among clinical isolates, and to infer their potential source when applied to large surveillance and reference datasets. This was applied to the current study data and the full PubMLST database for *Campylobacter*. Clinical isolates were assigned to single-linkage clusters (SLC), differing at 5 or fewer cgMLST loci (Cjc_cgc_5) from at least one other member of that group, as implemented automatically in BIGSdb (64).

Among 9,442 UK clinical isolates from 1997 to 2018 (8,524 *C. jejuni*, 918 *C. coli*), including the current study 4,823 (51%) isolates belonged to one of 954 SLC, containing between 2 and 192 members. A total of 71 SLC included 10 or more isolates. All were among *C. jejuni* isolates rather than *C. coli* and they occurred across 14 of the 44 *C. jejuni* clonal complexes. The duration and maximum allelic distance across these clusters is graphed in Figure 37, restricted to the human isolates from the current study (upper panel) and including all isolates on the PubMLST database (lower panel). Clusters extend over up to 15 years, with greater genetic diversity among long-lived clusters.

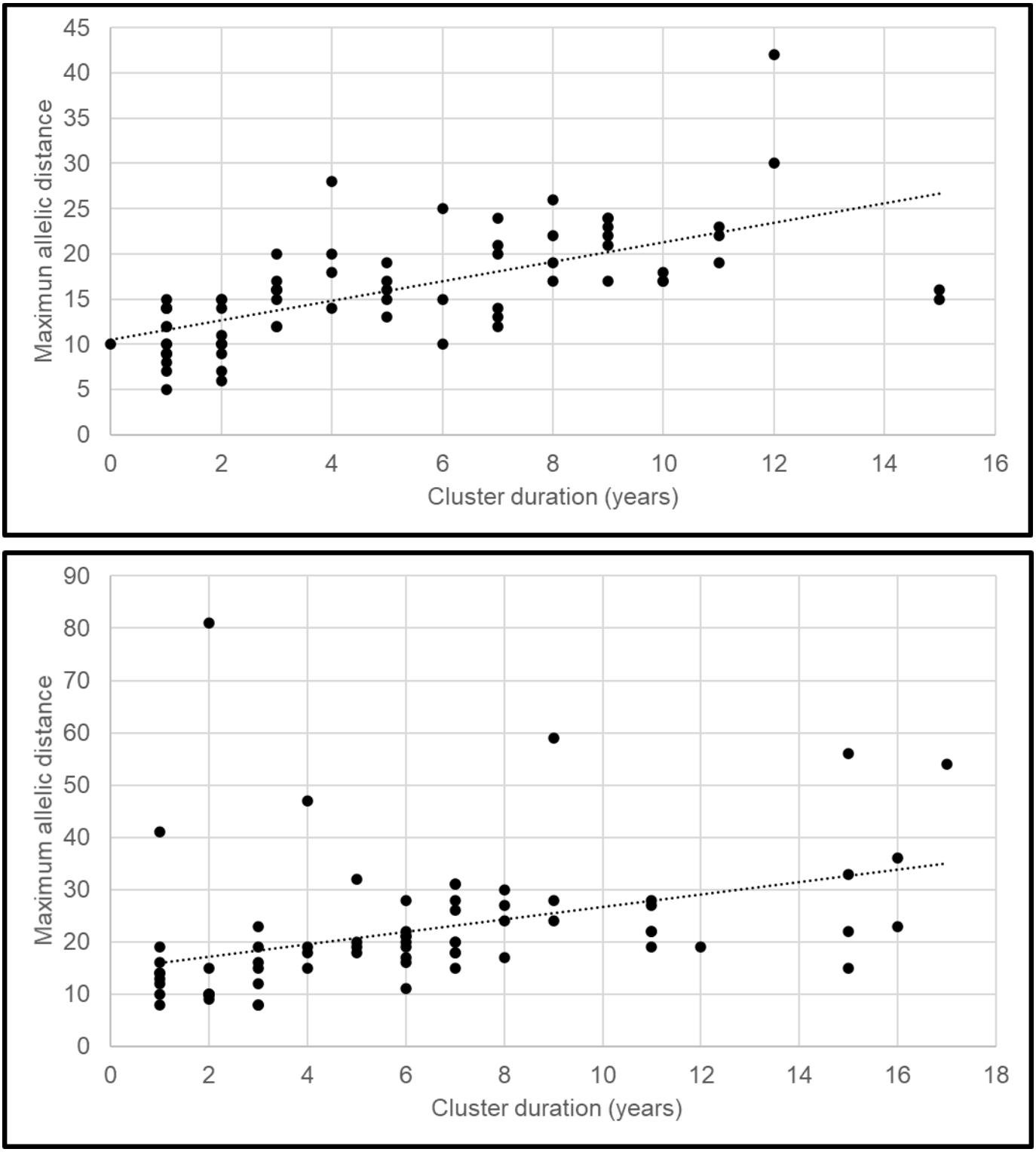



Figure 37. Relationship between maximum allelic distance and cluster duration for SLC containing ≥10 clinical isolates for UK clinical isolates 1997 – 2018 (upper panel) and incorporating global clinical and non-clinical sources on the PubMLST/campylobacter database (lower panel).

Individual clusters included isolates from both study sites. Analysing these 71 clusters, including scanning the full PubMLST database for non-UK human and for non-human animal sequences using the same 5 locus mismatch rule showed that many included isolates from chicken sources, some from ruminant, and some were multi-host. This may reflect differential sampling, at least in part, with these as the most extensively sampled non-human animals. The largest clusters mainly comprised human and chicken origin isolates (Figure 38).

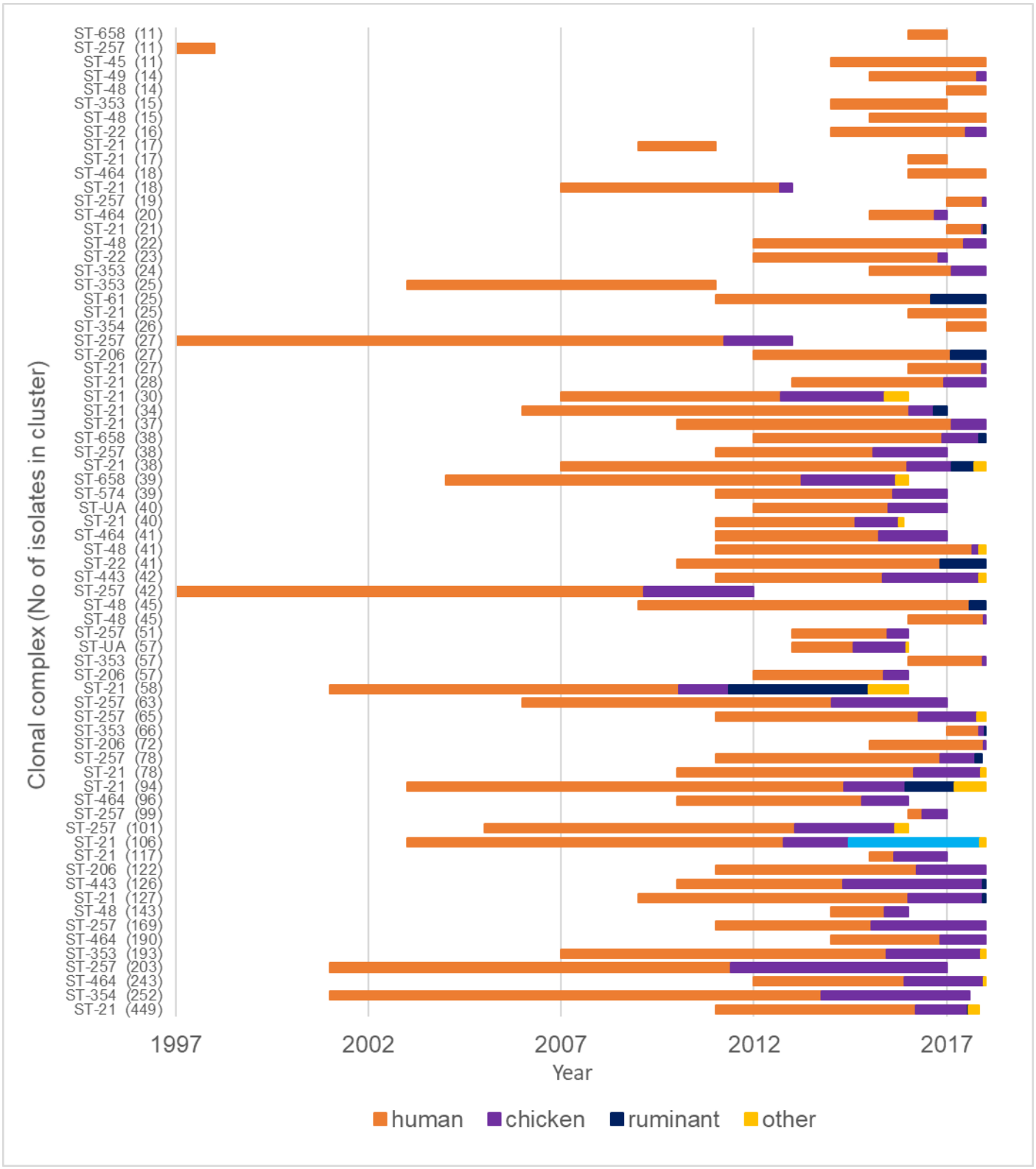



Figure 38. Timing and duration of single linkage clusters including all isolates (locations, time periods and sources) in the PubMLST database with proportions from humans (red), chickens (yellow), ruminants (blue) and other sources (grey) shown by colour. Cluster size ranges from 11 (top) to 499 (bottom).

As is evident from these figures this criterion identifies persistent lineages present over relatively long time periods as clusters, rather than being restricted to typical outbreaks clustered in time and space. The observed clusters thus likely represent a mixed epidemiology of some more typical outbreaks and some persistent lineages not meeting the usual epidemiological interpretation of an outbreak. However, some clusters are restricted to more highly related isolates. Those with maximum allelic distances of 5 are consistent with variation in repeat sampling from a single patient. These may represent strong epidemiological links with an example in Figure 39.

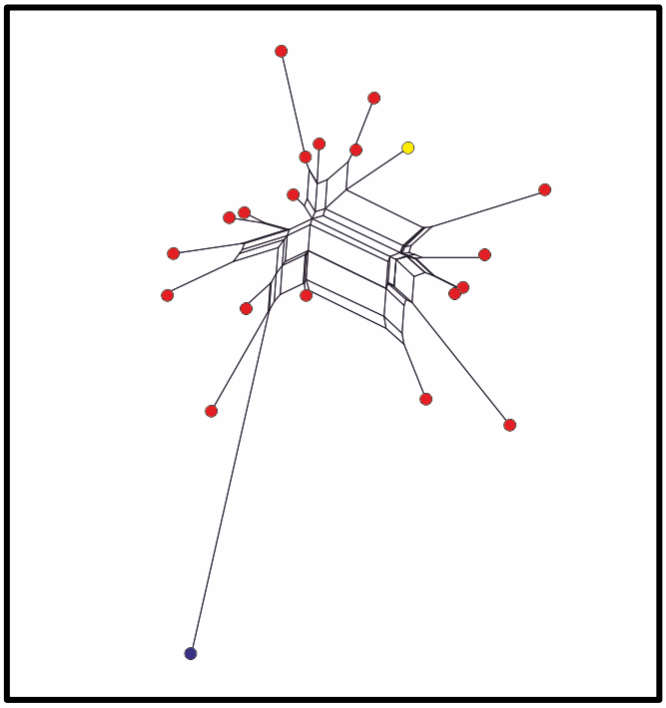


Figure 39. Genome Comparator analysis of an example cluster, SLC20233, with isolates belonging to ST-50 in ST-21 clonal complex and visualised in SplitsTree. This demonstrates a short allelic distance between human isolates (red), a chicken isolate (yellow) and a more distantly related ruminant isolate (blue).

## 9.3 Conclusions on observed clustering

Among *C. coli* isolates, genetic clustering is uncommon with relatively few and small clusters suggesting that most cases may be sporadic rather than linked as part of outbreaks, although with the caveat that multi-strain outbreaks are possible. In *C. jejuni* there is substantial genomic clustering at the 5 gene difference level of core genome MLST when analysed using single linkage clustering. However, many of these clusters are long lived, relatively diverse, and appear to

represent genomically stable lineages rather than providing evidence that cases have a close epidemiological relationship. Lineages with strong associations with just one among the various possible non-human sources may share a sustained ongoing source of infection. This is most evident for clusters including human and chicken isolates. It may support targeted intervention against specific sources and transmission routes even if not a typical outbreak. This is of potential public health importance given the large proportion of *C. jejuni* isolates that belong to clusters defined at this level of relatedness. These analyses also make it clear that further work is needed to better identify outbreaks and that the single linkage clustering that has proved effective in routine public health practice for *Salmonella* may not be optimal for *Campylobacter*.

## 10. Joint analyses of genomic attribution and epidemiology data

Genomic source attribution estimates the likely source of human infection but does not alone allow inference on the pathway. For example, isolates from a chicken source could conceivably lead to infection through cross contamination of other foodstuffs or even directly through the environment (65). This sentinel surveillance study did not include control data so that the types of formal integrated analysis possible in case control studies with genomic sequencing (65, 66) are not a possible or an intended part of the work.  Moreover, even though population level proportional attribution is possible as presented in section six, for individual isolates there is often substantial uncertainty as to the origin of infection given the presence of multi-host sequence types. However, the sentinel types identified as a complementary approach to attribution are likely to have a high probability of originating from that animal source. Comparison of exposure histories across these types, albeit a minor subset of the data human surveillance data, can contribute to both confirming their validity as sentinel types, and to identifying evidence for pathways of infection.

Among cases with both 1) a sentinel type strongly supportive of a particular reservoir source and 2) a complete exposure history, exposures were analysed across different sentinel types using the case-case analysis approach (67). This showed that those with ruminant types are significantly more likely to report consumption of lamb’s liver in the previous five days (odds ratio 15 [5-44]) (

Table 40) and less likely to report consumption chicken (odds ratio 0.3 [0.2-0.6]) (
Table 41). Moreover, 23% of those with a ruminant type reported having eaten lamb's liver. Those having a chicken type were more likely to have consumed chicken (78%) than those with a ruminant type, 51% of whom reported having eaten chicken and were very unlikely to report having eaten lamb's liver (2%).

Table 40. Association between consumption of lamb's liver and infection with sentinel types of *Campylobacter* (p-value <0.0001).

| **Type** | **Ate lamb's liver** | **%** | **Didn't eat lamb's liver** | **%** | **Total** | **Odds Ratio** | **95% CI** |
|---|---|---|---|---|---|---|---|
| Chicken | 6 | 2% | 300 | 98% | 306 | 1 | - |
| Ruminant | 10 | 23% | 33 | 77% | 43 | 15.15 | 5.18-44.35 |
| Unknown | 29 | 4% | 706 | 96% | 735 | 2.05 | 0.84-5.00 |

Table 41. Association between consumption of chicken and infection with sentinel types of *Campylobacter* (p-value <0.0001).

| **Type** | **Ate chicken** | **%** | **Didn't eat chicken** | **%** | **Total** | **Odds Ratio** | **95% CI** |
|---|---|---|---|---|---|---|---|
| Chicken | 286 | 78% | 81 | 22% | 367 | 1 | - |
| Ruminant | 22 | 51% | 21 | 49% | 43 | 0.3 | 0.16-0.56 |
| Unknown | 677 | 77% | 202 | 23% | 879 | 0.95 | 0.71-1.27 |

The concordance between the predictions based on genomic attribution and self-reported consumption supports chicken consumption as a cause of infection with chicken types and lamb's liver consumption as a cause of infection with ruminant types. This is evidence for a direct pathway to infection by eating meat contaminated with these host specific *Campylobacters* although not implying that this is the only route of transmission. There was no significant evidence for consumption of milk or other foods contributing to the risk of infection with ruminant types. The strong association of lamb's liver with ruminant type infection and 23% of those with these types recalling such consumption suggests that this may be an important cause of human infection, potentially causing several thousand cases of human infection in England each year.

# 11. Discussion, future work, and recommendations

This sentinel surveillance project estimated the proportion of *Campylobacter* attributable to chickens and to other animal species in England, described the pattern of antimicrobial resistance over time and across the food chain, evaluated evidence for clustering, and established a virtual and laboratory archive for further work that will support a range of other analyses to address other scientific questions.

## 11.1 Sources of human infection

Review of genetic attribution highlighted both the consistent finding of a large, and typically majority, contribution of the poultry reservoir to human infection and substantial gaps in approaches to validation and bias adjustment based on results in the existing literature (38). Validation showed greater reliability with the Structure algorithm than the asymmetric island model but also evidence for systematic bias in population genetic attribution using either model. Adjustment for bias in analysis using the Structure model increased the proportion of *C. jejuni* infections attributed to chicken and the proportion of *C. coli* attributed to ruminant sources. The percentage of *C. jejuni* attributed to chicken by this approach was 64% across the three years of the main study (October 2015 to September 2018), with most of the rest being attributed to ruminant sources. Analysis across the globally unique resource of fifteen years continuous genetic surveillance in Oxfordshire showed relatively stable attribution to each source across that time. The use of sentinel types and extrapolation allowed an independent approach to genetic source attribution for *C. jejuni*. This estimated 76% of human infection across the three years of the main study as originating from the chicken source. Both estimates are thus consistent with a substantial majority of human infection, in the order of two thirds to three quarters, coming from the chicken sources. The estimate, at 12%, for infection of ruminant origin from sentinel type analysis is substantially lower than by population genetic analysis using Structure in this study and in past published work (38) (supplementary table) with few exceptions (25, 68-70). Such a discrepancy would be expected if *Campylobacter* lineages restricted to ruminants among non-human sources transfer to humans less well than do *Campylobacter* lineages found in both ruminant and other animal species. In this case the sentinel type approach would be overly conservative. However, there is no evidence for, or against, this theory and it is not possible to test it in these data. Alternatively, given the restriction to three possible sources in the analysis using Structure, it is possible that infection coming from other sources, not considered in the analysis are being

misattributed to the ruminant source and falsely inflating the proportion apparently coming from this source. The sentinel approach, allowing a proportion to remain unknown does not suffer from this potential problem.

A further striking feature of the sentinel analysis results was the substantially higher representation among human isolates of sequence types typical of ruminants and present in both cattle and sheep, compared to those types present in cattle but not in sheep. Again, this contrasts with most published population genetic attribution results where the cattle reservoir is often identified as an important source of human infection when ruminant species are analysed separately (38) (supplementary table). In this study we merged ruminant species given the weak ability of current population genetic attribution approaches to determine the species of origin at this level but had the opportunity to consider cattle alone using sentinel types. Several mechanisms might explain this including that sheep but not cattle were in fact an important source of human infection in our study population, or that those types that are shared across cattle and sheep have greater capacity to transmit to humans than cattle restricted types. The current study thus raises uncertainties and the importance of further work to understand the role of ruminants, and specific ruminant species in the aetiology of human infection.

Joint analysis of sentinel types and exposure data, showing concordance in the expected origin of a type and the exposure histories of those with those types, supported the use of these sentinel types as valid indicators of the likely origin of these human isolates in the study population. The results of these analyses also showed that foodborne transmission (reported eating of chicken meat and ruminant offal) is an important pathway for infection as would be expected given the repeated identification of consumption of chicken in particular as an important risk factor for disease. Of note, for ruminant origin infection, indexed by infection with a ruminant sentinel type, only lamb's liver consumption was identified as a risk factor and not milk, beef or ox or calf liver, and fully 23% of those with a sentinel type suggestive of ruminant infection reported having eaten lamb's liver in the five days before illness. This finding that a large burden of human infection may be associated with the consumption of lamb's liver is novel. Even a conservative estimate, based on 12% of human disease coming from a ruminant source predicts that over 2% of human campylobacteriosis is associated with lamb's liver consumption, several thousand cases annually in the UK. Although there was no evidence for cattle specific types, or the consumption of milk, contributing substantially to human disease in the current study milk and milk products remain a potentially important source of human infection (71).

Overall, the current study thus demonstrates the high and sustained contribution of poultry to human *Campylobacter* infection England, consistent with and extending past literature, and raises substantial insights and questions into the scale of, and specific routes for, ruminant origin infection.

## 11.2 Methodology to determine the aetiology of human infection

This work employed adjustment for bias identified in population genetic approaches to monitoring the sources of human infection. As whole genome data becomes more available it is likely that accuracy of population genetic attribution will improve. However, bias may well remain an issue. The systematic review of published work on the population genetic attribution of human *Campylobacter* infection to source to date showed a complete absence of adjustment for such bias in the published literature, including in our own past publications. The current work therefore provides an important lesson for future approaches being developed on whole genome data on the importance of validation to estimate bias and performing adjustment or sensitivity analysis based on these findings.

Literature reviews during the course of the project did not identify an appropriate method to undertake attribution more accurately using whole genome data. In collaboration with Prof Nigel French, and Dr Jonathan Marshall in Massey University, New Zealand and Dr Simon Spencer at the University of Warwick, the genomic data from the current study and the virtual whole genome sequence archive curated as part of this project are being used to pilot the Massey group's approach of applying the asymmetric island model to whole genome data. Preliminary results have not identified substantial performance improvements over traditional seven gene MLST. Since completion of the project work a proposal for an efficient data mining approach to using whole genome data for attribution has been published (72). However, this does not identify substantial improvement over existing approaches using seven gene MLST either.

This therefore remains an area for further work before whole genome data can be used to substantially improve the accuracy of genomic attribution. The lack of marked improvements from population genetic analyses of these more extensive data using approaches tried to date also raises questions on the extent of additional potential benefit available, and whether this may be limited.

The current project also identified sentinel types across host reservoirs and used these in analysis to estimate the contribution of different sources to human infection as described above. This supported the estimates from adjusted population genetic attribution that a large proportion of

human infection continues to originate in the chicken reservoir. As larger whole genome sequenced reference datasets are developed this approach may grow more powerful if these data allow more accurate and larger scale identification of sentinel types and the stability, or otherwise, of their association with particular reservoir species.

The case-case comparison approach (67), contrasting risk factors across cases of infection with sentinel types strongly associated these particular food animals, proved informative in this study. As genomic surveillance becomes more prevalent this approach, combining genomic and epidemiological surveillance of cases in the absence of control data, may be a powerful method of identifying the pathways leading to human infection. There are also benefits to having population control data and their joint analysis with genomic surveillance data (65, 66). However, this is a much more major undertaking and may not be feasible as part of ongoing disease monitoring and investigation. Alternative approaches such as rolling population surveys in the Netherlands that are more sustainable might complement sentinel surveillance data (73) even though such approaches are not a direct substitute for controls (52, 73).

### 11.3 Antimicrobial resistance and use

The findings on antimicrobial resistance are not new, with for example high levels of fluoroquinolone resistance well established in human infection. However, the scale of this study and the demonstration of substantially lower levels in the earlier human isolates (1997-8) for many antibiotics provides a robust demonstration of the emergence of this resistance from low levels at that time, and assays genomic mechanisms involved on a large sample. Comparison across food animals strongly supports chicken as a particularly important source for fluoroquinolone resistant *Campylobacter* infection in England. Analysis within the project will continue through an ongoing doctoral research project at the University of Oxford.

Given national guidance advising antibiotic treatment of campylobacteriosis under only limited circumstances the level of treatment was reported was relatively high, and moreover varied markedly across the two sentinel sites. Although antibiotic treatment of human campylobacteriosis is not likely to be important in generating resistance in *Campylobacter*, with humans in England usually a dead-end host rather than a reservoir, it may be important in generating resistance in other bacteria present in those treated and therefore contributing to antimicrobial resistance more generally. The high levels of treatment might be explained by diagnosed cases being more severe and persistent and diagnostic testing potentially more common in vulnerable groups. However, the substantial variation in levels of antibiotic treatment by study site, and the higher levels of

treatment in non-elderly adults compared to children or the elderly infection highlight this as a potential area for improved antimicrobial stewardship.

## 11.4 Clustering and outbreaks

The data showed strong evidence for genomic clustering of *C. jejuni* but not *C. coli*. The meaning of this clustering is less clear. Large and sustained “clusters” of closely related isolates were identified by single linkage approaches using a threshold of up to 5 SNPs or 5 locus differences on cgMLST. Isolates from more than one non-human animal species were also present in some clusters. This is evidence for the persistence and substantial spread of relatively narrow clonal lineages, although with single linkage analysis there is some heterogeneity in some of these clusters. It argues against the interpretation of this level of relatedness, estimated using single linkage approaches, as strong evidence for an outbreak or shared epidemiological source. The availability of extensive animal and environmental isolates collections for analysis in the current study allowed this identification of the wide distribution of these clusters and the necessary caution in interpreting single linkage clustering, even at this 5 SNP or 5 locus thresholds for similarity.
In contrast some clusters were narrower across one or more of the dimensions of: genomic relatedness, range of non-human animal isolates, and time. These appeared more typical of point source or continuing source outbreaks (Figure 38) This is supported by recent findings in Denmark using average linkage analysis (74) and identifying substantial clustering, consistent with outbreaks, among apparently sporadic *C. jejuni* isolates.
Although single linkage approaches have worked well in the implementation of outbreak detection by Public Health England for *Salmonella*, this work suggests that other approaches may be more appropriate for *Campylobacter* and this is an area for development. Extensive core genome MLST sequenced datasets of human and non-human origin isolates provide a substantial resource to support this and sharing of these data effectively across the scientific and public health communities is essential to their effective research and public health applications.

## 11.5 Genome wide association and other basic science studies

The data obtained in this study support wider analysis and scientific benefit than envisaged in the study aims. As an example, analysis is underway with collaborators to test for evidence of any genetic associations with clinical outcome in this dataset using a genome wide association study approach. This involves testing for associations between patient reported measures of disease

severity and bacterial genetic variation.  This will continue, supported by the NIHR Health Protection Research Unit in Gastrointestinal Infection. More broadly the project data is being made accessible through PubMLST and the study investigators will continue to collaborate with others wishing to use these data.

## 11.6 Future sentinel surveillance

Isolates from the Oxfordshire site are still being archived using university resource to maintain this globally unique longitudinal dataset but without current sustainable funding for this work. The North East site sampling has stopped. Other countries including Ireland and Denmark have started sentinel surveillance activity. This is being developed using resources developed and made accessible through PubMLST. Public Health England continues to collect some isolates from the Midlands as part of routing public health activity. The unique historical resource including continuity of sampling from past Defra and Food Standards Agency investments, alongside substantial food and food animal sampling offers a strong foundation to allow the UK to continue to make a leading global contribution in this area.

## 11.7 Recommendations

### Using information reported on the sources of human infection to support control

1. This work reaffirms the high and consistent contribution of poultry to the burden of *Campylobacter* in humans and supports ongoing prioritisation of this area for control.
2. The potential importance of the specific risk factor of consumption of lamb's liver merits further investigation to support interventions that may include guidance on handling and cooking.

### Maximising accuracy and use of genomic attribution data

3. Potential biases in genomic attribution should be estimated and appropriate adjustment or sensitivity analysis performed. This has not been standard practice in the past and is important to improve accuracy as demonstrated in this work.
4. Genomic attribution data should be analysed jointly with wider data such as exposure data among those with campylobacteriosis, and ideally population or control estimates of these same exposures in the population. This can confirm and extend understanding of the sources of human infection inferred from genomic attribution alone.

5. Method development for attribution using whole genome data needs further work to allow use of the growing data available, and complementary methods such as the sentinel type approach applied here should be trialled and evaluated to complement population genetic algorithm-based attribution.

### Antimicrobial use and resistance

6. The potential gaps and variability in antimicrobial stewardship highlighted in this study should guide review of opportunities for improved antimicrobial stewardship in the management of gastroenteritis.
7. The high levels of antimicrobial resistance across several classed of antibiotics, and generally high but varying levels of resistance across food animal species, emphasise the importance of *Campylobacter* as a model organism for a One Health approach to antimicrobial resistance and to monitor the impact of interventions in the agricultural sector on the transmission of resistance determinants to humans.

### Integration and data access to optimise usefulness of investment

8. Inferences on outbreaks, sources of human infection, and the ecology of antibiotic resistance should be based on both combined contemporaneous human case and food animal sampling, and joint analysis of these with large and wide-ranging datasets available through global collaboration and collation of data. This supports better interpretation of clustering and outbreak investigation, overall disease attribution to source, and the study of antimicrobial resistance. Research commissioning can be a powerful tool to support this integrated and collaborative approach to this research and joint resource development.
9. Shared and accessible data is critical to maximise benefits from local and global work in this area. The approach of this project has been to make data available publicly as the study progresses and this is recommended to researchers and research commissioners.
10. Alongside sharing and access, the interface to support a range of users is critical to optimise the use of these complex and diverse data. The application of data visualisation approaches piloted in this study can support this wider use of data.

### Questions emerging from this work

11. The contribution of infection from the cattle and sheep reservoirs individually and collectively should be addressed in more detail alongside continued focus on the chicken source. Findings

such as the relative lack of human disease in this study due to sequence types found in cattle but not sheep or chicken raise questions that may open new areas of understanding.

12. Implementation research is needed to optimise the use of genomic surveillance in the detection and investigation of *Campylobacter* outbreaks. This study and emerging work in other countries is demonstrating both limitations and uses for these approaches in outbreak surveillance. These are now ready to be trialled within an evaluative framework to optimise their application and maximise their impact on disease control.

### A critical research and public health resource

13. The data from the current project and uninterrupted past longitudinal follow up in Oxfordshire is unique globally in terms of human surveillance and strongly complemented by FSA funded studies among food, and food animals in the UK. The use and maintenance of this unmatched resource is a priority for the UK and globally.

## 12. References


1. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut. 2012;61(1):69-77.
https://doi.org/10.1136/gut.2011.238386

2. Man SM. The clinical importance of emerging Campylobacter species. Nature reviews Gastroenterology & hepatology. 2011;8(12):669.
https://doi.org/10.1038/nrgastro.2011.191

3. Liu F, Ma R, Wang Y, Zhang L. The clinical importance of Campylobacter concisus and other human hosted Campylobacter species. Frontiers in cellular and infection microbiology. 2018;8:243.
https://doi.org/10.3389/fcimb.2018.00243

4. Cody AJ, McCarthy NM, Wimalarathna HL, Colles FM, Clark L, Bowler IC, et al. A longitudinal 6-year study of the molecular epidemiology of clinical Campylobacter isolates in Oxfordshire, United Kingdom. Journal of clinical microbiology. 2012;50(10):3193-201.
https://doi.org/10.1128/jcm.01086-12

5. Sahin O, Yaeger M, Wu Z, Zhang Q. Campylobacter-associated diseases in animals. Annual review of animal biosciences. 2017;5:21-42.
https://doi.org/10.1146/annurev-animal-022516-022826

6. Reuter M, Mallett A, Pearson BM, van Vliet AH. Biofilm formation by Campylobacter jejuni is increased under aerobic conditions. Appl Environ Microbiol. 2010;76(7):2122-8.
https://doi.org/10.1128/aem.01878-09

7. Snelling WJ, Moore JE, McKenna JP, Lecky DM, Dooley JS. Bacterial-protozoa interactions; an update on the role these phenomena play towards human illness. Microbes Infect. 2006;8(2):578-87.
https://doi.org/10.1016/j.micinf.2005.09.001

8. Bronowski C, James CE, Winstanley C. Role of environmental survival in transmission of Campylobacter jejuni. FEMS microbiology letters. 2014;356(1):8-19.
https://doi.org/10.1111/1574-6968.12488

9. Bolton DJ. Campylobacter virulence and survival factors. Food microbiology. 2015;48:99-108. https://doi.org/10.1016/j.fm.2014.11.017

10. Moore JE, Corcoran D, Dooley JS, Fanning S, Lucey B, Matsuda M, et al. Campylobacter. Veterinary research. 2005;36(3):351-82.

11. Gibbens JC, Pascoe SJ, Evans SJ, Davies RH, Sayers AR. A trial of biosecurity as a means to control Campylobacter infection of broiler chickens. Preventative Veterinary Medicine. 2001;48(2):85-99.
https://doi.org/10.1016/s0167-5877(00)00189-6

12. Dhama K, Rajagunalan S, Chakraborty S, Verma AK, Kumar A, Tiwari R, et al. Food-borne pathogens of animal origin-diagnosis, prevention, control and their zoonotic significance: a review. Pak J Biol Sci. 2013;16(20):1076-85.
https://doi.org/10.3923/pjbs.2013.1076.1085

13. Teunis PFM, Bonacic Marinovic A, Tribble DR, Porter CK, Swart A. Acute illness from Campylobacter jejuni may require high doses while infection occurs at low doses. Epidemics. 2018;24:1-20.
https://doi.org/10.1016/j.epidem.2018.02.001

14. Awofisayo-Okuyelu A, Hall I, Adak G, Hawker JI, Abbott S, Mc CN. A systematic review and meta-analysis on the incubation period of Campylobacteriosis. Epidemiol Infect. 2017;145(11):2241-53.
https://doi.org/10.1017/s0950268817001303

15. Hu L, Kopecko DD. Campylobacter species. Food Safety: Apple Academic Press; 2018. p. 73-110.

16. Allos BM. Association between Campylobacter infection and Guillain Barré syndrome. J Infect Dis. 1997;176(Suppl.2):S125-8.
https://doi.org/10.1086/513783

17. Platts-Mills JA, Kosek M. Update on the burden of Campylobacter in developing countries. Curr Opin Infect Dis. 2014;27(5):444-50.
https://doi.org/10.1097/qco.0000000000000091

18. Devleesschauwer B, Bouwknegt M, Mangen M-JJ, Havelaar AH. Health and economic

burden of Campylobacter. Campylobacter: Elsevier; 2017. p. 27-40.
https://doi.org/10.1016/b978-0-12-803623-5.00002-2

19. Jolley KA, Bray JE, Maiden MC. Open-access bacterial population genomics: BIGSdb software, the PubMLST. org website and their applications. Wellcome open research. 2018;3.
https://doi.org/10.12688/wellcomeopenres.14826.1

20. Dingle KE, Colles FM, Ure R, Wagenaar JA, Duim B, Bolton FJ, et al. Molecular characterization of Campylobacter jejuni clones: a basis for epidemiologic investigation. Emerging infectious diseases. 2002;8(9):949.
https://doi.org/10.3201/eid0809.02-0122

21. Cody AJ, Bray JE, Jolley KA, McCarthy ND, Maiden MCJ. Core Genome Multilocus Sequence Typing Scheme for Stable, Comparative Analyses of Campylobacter jejuni and C. coli Human Disease Isolates. J Clin Microbiol. 2017;55(7):2086-97.
https://doi.org/10.1128/jcm.00080-17

22. Gundogdu O, Bentley SD, Holden MT, Parkhill J, Dorrell N, Wren BW. Re-annotation and re-analysis of the Campylobacter jejuni NCTC11168 genome sequence. BMC Genomics. 2007;8:162.
https://doi.org/10.1186/1471-2164-8-162

23. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nature Reviews Microbiology. 2013;11(10):728-36.
https://doi.org/10.1038/nrmicro3093

24. Sheppard SK, Dallas JF, Strachan NJ, MacRae M, McCarthy ND, Wilson DJ, et al. Campylobacter genotyping to determine the source of human infection. Clinical Infectious Diseases. 2009;48(8):1072-8.
https://doi.org/10.1086/597402

25. Kittl S, Heckel G, Korczak BM, Kuhnert P. Source attribution of human Campylobacter isolates by MLST and fla-typing and association of genotypes with quinolone resistance. PLoS One. 2013;8(11):e81796.
https://doi.org/10.1371/journal.pone.0081796

26. Stanley K, Jones K. Cattle and sheep farms as reservoirs of Campylobacter. Journal of

applied microbiology. 2003;94:104-13.
https://doi.org/10.1046/j.1365-2672.94.s1.12.x

27. Shange N, Gouws P, Hoffman LC. Campylobacter and Arcobacter species in food-producing animals: prevalence at primary production and during slaughter. World Journal of Microbiology and Biotechnology. 2019;35(9):146.
https://doi.org/10.1007/s11274-019-2722-x

28. Colles FM, Ali JS, Sheppard SK, McCarthy ND, Maiden MC. Campylobacter populations in wild and domesticated Mallard ducks (Anas platyrhynchos). Environ Microbiol Rep. 2011;3(5):574-80.
https://doi.org/10.1111/j.1758-2229.2011.00265.x

29. Griekspoor P, Colles FM, McCarthy ND, Hansbro PM, Ashhurst-Smith C, Olsen B, et al. Marked host specificity and lack of phylogeographic population structure of Campylobacter jejuni in wild birds. Mol Ecol. 2013;22(5):1463-72.
https://doi.org/10.1111/mec.12144

30. Cody AJ, McCarthy ND, Bray JE, Wimalarathna HM, Colles FM, Jansen van Rensburg MJ, et al. Wild bird-associated Campylobacter jejuni isolates are a consistent source of human disease, in Oxfordshire, United Kingdom. Environ Microbiol Rep. 2015;7(5):782-8.
https://doi.org/10.1111/1758-2229.12314

31. Khan IU, Hill S, Nowak E, Edge TA. Effect of incubation temperature on the detection of thermophilic Campylobacter species from freshwater beaches, nearby wastewater effluents, and bird fecal droppings. Appl Environ Microbiol. 2013;79(24):7639-45.
https://doi.org/10.1128/aem.02324-13

32. Bolton F, Surman S, Martin K, Wareing D, Humphrey T. Presence of Campylobacter and Salmonella in sand from bathing beaches. Epidemiology & Infection. 1999;122(1):7-13.
https://doi.org/10.1017/s0950268898001915

33. Colles F, Jones K, Harding R, Maiden M. Genetic diversity of Campylobacter jejuni isolates from farm animals and the farm environment. Appl Environ Microbiol. 2003;69(12):7409-13.
https://doi.org/10.1128/aem.69.12.7409-7413.2003

34. Pitkänen T. Review of Campylobacter spp. in drinking and environmental waters. Journal of microbiological methods. 2013;95(1):39-47.
https://doi.org/10.1016/j.mimet.2013.06.008

35. Jones K, Obiri-Danso K. Non-compliance of beaches with the EU directives of bathing water quality: evidence of non-point sources of pollution in Morecambe Bay. Journal of applied microbiology. 1998;85(S1):101S-7S.
https://doi.org/10.1111/j.1365-2672.1998.tb05288.x

36. Colles FM, McCarthy ND, Howe JC, Devereux CL, Gosler AG, Maiden MC. Dynamics of Campylobacter colonization of a natural host, Sturnus vulgaris (European starling). Environmental Microbiology. 2009;11(1):258-67.
https://doi.org/10.1111/j.1462-2920.2008.01773.x

37. EFSA. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2013. EFSA Journal 2015. 2015;13(1).
https://doi.org/10.2903/j.efsa.2015.3991

38. Cody AJ, Maiden MC, Strachan NJ, McCarthy ND. A systematic review of source attribution of human campylobacteriosis using multilocus sequence typing. Euro Surveill. 2019;24(43).
https://doi.org/10.2807/1560-7917.es.2019.24.43.1800696

39. Nylen G, Dunstan F, Palmer SR, Andersson Y, Bager F, Cowden J, et al. The seasonal distribution of campylobacter infection in nine European countries and New Zealand. Epidemiology and Infection. 2002;128(3):383-90.
https://doi.org/10.1017/s0950268802006830

40. Louis VR, Gillespie IA, O'Brien SJ, Russek-Cohen E, Pearson AD, Colwell RR. Temperature-driven Campylobacter seasonality in England and Wales. Appl Environ Microbiol. 2005;71(1):85-92.
https://doi.org/10.1128/aem.71.1.85-92.2005

41. Wallace JS, Stanley KN, Currie JE, Diggle PJ, Jones K. Seasonality of thermophilic Campylobacter populations in chickens. Journal of Applied Microbiology. 1997;82:219-24.
https://doi.org/10.1111/j.1365-2672.1997.tb03576.x

42. Kramer JM, Frost JA, Bolton FJ, Wareing DR. Campylobacter contamination of raw meat and poultry at retail sale: identification of multiple types and comparison with isolates from human infection. Journal of Food Protection. 2000;63(12):1654-9.
https://doi.org/10.4315/0362-028x-63.12.1654

43. Niederer L, Kuhnert P, Egger R, Buttner S, Hachler H, Korczak BM. Genotypes and Antibiotic Resistances of Campylobacter jejuni and Campylobacter coli Isolates from Domestic and Travel-Associated Human Cases. Appl Environ Microbiol. 2012;78(1):288-91. https://doi.org/10.1128/aem.06194-11

44. Moore JE, Barton MD, Blair IS, Corcoran D, Dooley JS, Fanning S, et al. The epidemiology of antibiotic resistance in Campylobacter. Microbes and infection. 2006;8(7):1955-66.

45. Kapperud G, Espeland G, Wahl E, Walde A, Herikstad H, Gustavsen S, et al. Factors associated with increased and decreased risk of Campylobacter infection: a prospective case-control study in Norway. Am J Epidemiol. 2003;158(3):234-42. https://doi.org/10.1093/aje/kwg139

46. Skirrow MB, Blaser MJ. Clinical aspects of Campylobacter infection. In: Nachamkin I, Blaser MJ, editors. Campylobacter. Washington, D.C.: ASM Press; 2000. p. 69-88.

47. Engberg J, Aarestrup FM, Taylor DE, Gerner-Smidt P, Nachamkin I. Quinolone and macrolide resistance in Campylobacter jejuni and C. coli: resistance mechanisms and trends in human isolates. Emerging Infectious Diseases. 2001;7(1):24-34. https://doi.org/10.3201/eid0701.010104

48. Nachamkin I, Ung H, Li M. Increasing Fluoroquinolone Resistance in Campylobacter jejuni, Pennsylvania, USA, 1982–20011. Emerging infectious diseases. 2002;8(12):1501. https://doi.org/10.3201/eid0812.020115

49. D'Lima CB, Miller WG, Mandrell RE, Wright SL, Siletzky RM, Carver DK, et al. Clonal population structure and specific genotypes of multidrug-resistant Campylobacter coli from turkeys. Appl Environ Microbiol. 2007;73(7):2156-64. https://doi.org/10.1128/aem.02346-06

50. Szczepanska B, Andrzejewska M, Spica D, Klawe JJ. Prevalence and antimicrobial resistance of Campylobacter jejuni and Campylobacter coli isolated from children and environmental sources in urban and suburban areas. BMC microbiology. 2017;17(1):80. https://doi.org/10.1186/s12866-017-0991-9

51. Wallace RL, Bulach DM, Jennison AV, Valcanis M, McLure A, Smith JJ, et al. Molecular characterization of Campylobacter spp. recovered from beef, chicken, lamb and pork products at retail in Australia. PLoS One. 2020;15(7):e0236889.
https://doi.org/10.1371/journal.pone.0236889

52. Inns T, Curtis D, Crook P, Vivancos R, Gardiner D, McCarthy N, et al. Are food exposures obtained through commercial market panels representative of the general population? Implications for outbreak investigations. Epidemiol Infect. 2019;147:e99.
https://doi.org/10.1017/s0950268819000219

53. Smid JH, Mughini Gras L, de Boer AG, French NP, Havelaar AH, Wagenaar JA, et al. Practicalities of using non-local or non-recent multilocus sequence typing data for source attribution in space and time of human campylobacteriosis. PLoS One. 2013;8(2):e55029.
https://doi.org/10.1371/journal.pone.0055029

54. McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MC, et al. Host-associated genetic import in Campylobacter jejuni. Emerging Infectious Diseases. 2007;13(2):267-72.
https://doi.org/10.3201/eid1302.060620

55. Sheppard SK, Colles F, Richardson J, Cody AJ, Elson R, Lawson A, et al. Host Association of Campylobacter Genotypes Transcends Geographic Variation. Appl Environ Microbiol. 2010;76(15):5269-77.
https://doi.org/10.1128/aem.00124-10

56. Whyte P, McGill K, Cowley D, Madden RH, Moran L, Scates P, et al. Occurrence of Campylobacter in retail foods in Ireland. Int J Food Microbiol. 2004;95(2):111-8.
https://doi.org/10.1016/j.ijfoodmicro.2003.10.018

57. Di Giannatale E, Calistri P, Di Donato G, Decastelli L, Goffredo E, Adriano D, et al. Thermotolerant Campylobacter spp. in chicken and bovine meat in Italy: Prevalence, level of contamination and molecular characterization of isolates. PLoS One. 2019;14(12):e0225957.
https://doi.org/10.1371/journal.pone.0225957

58. Kwan PS, Birtles A, Bolton FJ, French NP, Robinson SE, Newbold LS, et al. Longitudinal study of the molecular epidemiology of Campylobacter jejuni in cattle on dairy farms. Appl

Environ Microbiol. 2008;74(12):3626-33.
https://doi.org/10.1128/aem.01669-07

59. French N, Barrigas M, Brown P, Ribiero P, Williams N, Leatherbarrow H, et al. Spatial epidemiology and natural population structure of Campylobacter jejuni colonizing a farmland ecosystem. Environmental Microbiology. 2005;7(8):1116-26.
https://doi.org/10.1111/j.1462-2920.2005.00782.x

60. Mickan L, Doyle R, Valcanis M, Dingle KE, Unicomb L, Lanser J. Multilocus sequence typing of Campylobacter jejuni isolates from New South Wales, Australia. Journal of Applied Microbiology. 2007;102(1):144-52.
https://doi.org/10.1111/j.1365-2672.2006.03049.x

61. Noormohamed A, Fakhr MK. Molecular Typing of Campylobacter jejuni and Campylobacter coli Isolated from Various Retail Meats by MLST and PFGE. Foods. 2014;3:82-93.
https://doi.org/10.3390/foods3010082

62. Painset A, Day M, Doumith M, Rigby J, Jenkins C, Grant K, et al. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Campylobacter jejuni and Campylobacter coli isolated from cases of diarrhoeal disease in England and Wales, 2015-16. J Antimicrob Chemother. 2020;75(4):883-9.
https://doi.org/10.1093/jac/dkz539

63. Cody AJ, Clarke L, Bowler IC, Dingle KE. Ciprofloxacin-resistant campylobacteriosis in the UK. Lancet. 2010;376(9757):1987.
https://doi.org/10.1016/s0140-6736(10)62261-1

64. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11(1):595.
https://doi.org/10.1186/1471-2105-11-595

65. Mughini Gras L, Smid JH, Wagenaar JA, de Boer AG, Havelaar AH, Friesema IH, et al. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One. 2012;7(8):e42599.
https://doi.org/10.1371/journal.pone.0042599

66. Mossong J, Mughini-Gras L, Penny C, Devaux A, Olinger C, Losch S, et al. Human Campylobacteriosis in Luxembourg, 2010-2013: A Case-Control Study Combined with Multilocus Sequence Typing for Source Attribution and Risk Factor Analysis. Scientific reports.

2016;6:20939.
https://doi.org/10.1038/srep20939

67. McCarthy N, Giesecke J. Case-case comparisons to study causation of common infectious diseases. Int J Epidemiol. 1999;28(4):764-8.
https://doi.org/10.1093/ije/28.4.764

68. Rosner BM, Schielke A, Didelot X, Kops F, Breidenbach J, Willrich N, et al. A combined case-control and molecular source attribution study of human Campylobacter infections in Germany, 2011-2014. Scientific reports. 2017;7(1):5139.
https://doi.org/10.1038/s41598-017-05227-x

69. Di Giannatale E, Garofolo G, Alessiani A, Di Donato G, Candeloro L, Vencia W, et al. Tracing Back Clinical Campylobacter jejuni in the Northwest of Italy and Assessing Their Potential Source. Front Microbiol. 2016;7:887.
https://doi.org/10.3389/fmicb.2016.00887

70. Mullner P, Jones G, Noble A, Spencer SE, Hathaway S, French NP. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Analysis. 2009;29(7):970-84.
https://doi.org/10.1111/j.1539-6924.2009.01224.x

71. Fernandes AM, Balasegaram S, Willis C, Wimalarathna HM, Maiden MC, McCarthy ND. Partial failure of milk pasteurization as a risk for the transmission of Campylobacter from cattle to humans. Clinical Infectious Diseases. 2015;61(6):903-9.
https://doi.org/10.1093/cid/civ431

72. Perez-Reche FJ, Rotariu O, Lopes BS, Forbes KJ, Strachan NJC. Mining whole genome sequence data to efficiently attribute individuals to source populations. Scientific reports. 2020;10(1):12124.
https://doi.org/10.1038/s41598-020-68740-6

73. Friesema IH, van Gageldonk-Lafeber AB, van Pelt W. Extension of traditional infectious disease surveillance with a repeated population survey. Eur J Public Health. 2015;25(1):130-4.
https://doi.org/10.1093/eurpub/cku122

74. Joensen KG, Kiil K, Gantzhorn MR, Nauerby B, Engberg J, Holt HM, et al. Whole-Genome Sequencing to Detect Numerous Campylobacter jejuni Outbreaks and Match Patient Isolates to

Sources, Denmark, 2015-2017. Emerg Infect Dis. 2020;26(3):523-32.
https://doi.org/10.3201/eid2603.190947

75. Cody AJ, McCarthy ND, Jansen van Rensburg M, Isinkaye T, Bentley S, Parkhill J, et al. Real-time genomic epidemiology of human Campylobacter isolates using whole genome multilocus sequence typing. J Clin Microbiol. 2013;51(8):2526-34.
https://doi.org/10.1128/jcm.00066-13

76. Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MC. A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative Neisseria meningitidis genomes. BMC Genomics. 2014;15:1138.
https://doi.org/10.1186/1471-2164-15-1138

77. Sheppard SK, Cheng L, Meric G, de Haan CP, Llarena AK, Marttinen P, et al. Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. Mol Ecol. 2014;23(10):2442-51.
https://doi.org/10.1111/mec.12742

78. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. Proc Natl Acad Sci U S A. 2013;110:11923-7.
https://doi.org/10.1073/pnas.1305559110

79. Zerbino D. Using the Velvet de novo Assembler for Short-Read Sequencing Technologies. Current Protocols in Bioinformatics. 2010;11(5):1-12.
https://doi.org/10.1002/0471250953.bi1105s31

80. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.
https://doi.org/10.1089/cmb.2012.0021

81. Cody AJ, McCarthy NM, Wimalarathna HL, Colles FM, Clark L, Bowler IC, et al. A longitudinal six-year study of the molecular epidemiology of clinical Campylobacter isolates in Oxfordshire, UK. J Clin Microbiol. 2012;50(10):3193-201.
https://doi.org/10.1128/jcm.01086-12

82. Mather AE, Vaughan TG, French NP. Molecular Approaches to Understanding Transmission and Source Attribution in Nontyphoidal Salmonella and Their Application in Africa. Clin Infect

Dis. 2015;61 Suppl 4:S259-65.
https://doi.org/10.1093/cid/civ727

83. McRobb E, Sarovich DS, Price EP, Kaestli M, Mayo M, Keim P, et al. Tracing melioidosis back to the source: using whole-genome sequencing to investigate an outbreak originating from a contaminated domestic water supply. J Clin Microbiol. Apr;. J Clin Microbiol. 2015;53(4):1144-8.
https://doi.org/10.1128/jcm.03453-14

84. Bergholz TM, Moreno Switt AI, Wiedmann M. Omics approaches in food safety: fulfilling the promise? . Trends Microbiol. 2014 22(5):275-81.
https://doi.org/10.1016/j.tim.2014.01.006

85. Gopinath G, Hari K, Jain R, Mammel MK, Kothary MH, Franco AA, et al. The Pathogen-annotated Tracking Resource Network (PATRN) system: a web-based resource to aid food safety, regulatory science, and investigations of foodborne pathogens and disease. . Food Microbiol. 2013 34(2):303-18.
https://doi.org/10.1016/j.fm.2013.01.001

86. Ragimbeau C, Colin S, Devaux A, Decruyenaere F, Cauchie HM, Losch S, et al. Investigating the host specificity of Campylobacter jejuni and Campylobacter coli by sequencing gyrase subunit A. BMC Microbiol. 2014;14:205.
https://doi.org/10.1186/s12866-014-0205-7

87. Boysen L, Rosenquist H, Larsson JT, Nielsen EM, Sorensen G, Nordentoft S, et al. Source attribution of human campylobacteriosis in Denmark. Epidemiology and Infection. 2014;142(8):1599-608.
https://doi.org/10.1017/s0950268813002719

88. Pascoe B, Meric G, Yahara K, Wimalarathna H, Murray S, Hitchings MD, et al. Local genes for local bacteria: Evidence of allopatry in the genomes of transatlantic Campylobacter populations. Mol Ecol. 2017;26(17):4497-508.
https://doi.org/10.7287/peerj.preprints.2638

89. Thepault A, Meric G, Rivoal K, Pascoe B, Mageiros L, Touzain F, et al. Genome-Wide Identification of Host-Segregating Epidemiological Markers for Source Attribution in Campylobacter jejuni. Appl Environ Microbiol. 2017;83(7).
https://doi.org/10.1128/aem.03085-16

90. Bacigalupe R, Lindsay D, Edwards G, Fitzgerald JR. Population Genomics of Legionella longbeachae and Hidden Complexities of Infection Source Attribution. Emerg Infect Dis. 2017;23(5):750-7.
https://doi.org/10.3201/eid2305.161165

91. Buultjens AH, Chua KYL, Baines SL, Kwong J, Gao W, Cutcher Z, et al. A Supervised Statistical Learning Approach for Accurate Legionella pneumophila Source Attribution during Outbreaks. Appl Environ Microbiol. 2017;83(21).
https://doi.org/10.1128/aem.01482-17

92. 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 Genom. 2017;3(10):e000135.
https://doi.org/10.1099/mgen.0.000135

93. Williamson DA, Baines SL, Carter GP, da Silva AG, Ren X, Sherwood J, et al. Genomic Insights into a Sustained National Outbreak of Yersinia pseudotuberculosis. Genome Biol Evol. 2016;8(12):3806-14.
https://doi.org/10.1093/gbe/evw285

# 13. Appendices

## 13.1 Appendix 1. Study group

### Report authors

University of Warwick

Noel McCarthy
Eileen Taylor

University of Oxford
Martin Maiden
Alison Cody
Melissa Jansen van Rensburg
Margaret Varga
Sophie Hedges

Public Health England
Craig Swift
Frieda Jorgensen
Tim Dallman
Anaïs Painset
Amy Douglas

## Additional study contributors

University of Oxford
Keith Jolley
James Bray
Dessislava Veltcheva

Public Health England
Kathie Grant
Nicola Elviss
Richard Elson
David van Santen
Clare Humphreys
Joan Henderson
Alison Waldram
Russell Gorton
Petra Manley
Animal and Plant Health Agency
John Rogers
Ana Vidal
Rob Davies

The role of Alison Cody to this project, the genetic epidemiology of *Campylobacter* more widely, and the community is particularly acknowledged. Sadly Alison died shortly after completing her substantial contributions to this project and is much missed by those involved in this work and the wider *Campylobacter* research and public health community.

## Acknowledgements

The Food Standards Agency funded this work and nominated external scientific experts and Food Standards Agency staff to the steering group as below.
External experts

Sarah O’Brien (Chair)
Julian Ketley
John Cowden
John Coia

Food Standards Agency
Kara Thomas
Kirsten Stone
Darren Holland
Paul Cook
Robin Clifford
Tara French
Adam Hardgrave

Food Standards Scotland (observer)
Keith Forbes

University of Aberdeen (observer)
Jane Horne

## 13.2 Appendix 2. Comparison of sequencing at PHE and the Sanger Institute

### Background and aim

Genome sequencing using Illumina High-Seq technology is considered highly accurate and reproducible (75, 76) and joint analysis of data processed at different sites is increasingly common (77, 78). Empirical tests of reproducibility among individual pathogens and centres are limited. The Public Health England (PHE) sequencing service offered improved timeliness over that available through the Sanger Institute. Confirmation of the comparability with Sanger Institute results was conducted as required by the Food Standards Agency to allow confidence in the quality of results obtained at PHE and when jointly analysing data this project alongside the extensive food and animal isolate collections that have been sequenced elsewhere, mainly at the Sanger Institute. The variation between methods at each site comprises the use of different library preparation protocols and genome assembly pipelines, with the actual sequencing at both sites using Illumina High-Seq 2000 sequencers.

### Methodology

Direct comparison
DNA from 22 Oxfordshire clinical *Campylobacter* isolates for which existing whole genome sequence (WGS) data had already been obtained from the Sanger Institute were sent to PHE for sequencing. These were isolates from four patients who supplied two samples during the course of their illness (all eight samples sequenced at PHE), two patients with single samples sequenced once (both sequenced at PHE), and seven patients with single isolates from whom duplicate DNA preparations had been sequenced in different sequencing runs at the Sanger Institute (with both DNA preparations available and re-sequenced for five patients and one preparation from each of the other two). The chosen samples covered a broad range of clonally distant *Campylobacter* clonal complexes.
Sequencing used Illumina High-Seq 2000 technology. Library preparation at the Sanger Institute used Covaris DNA shearing and at PHE Nextera fragmentation and tagging. Short-read data from the Sanger Institute were assembled using Velvet version 1.2.01 (79) shuffle and optimisation scripts in the BIGSdb (64) bioinformatics pipeline at Oxford. PHE sequenced isolates were assembled using the SPAdes genome assembler (80) in the PHE pipeline.
Contigs for each isolate were uploaded to the pubmlst.org/campylobacter database, which automatically identified loci, tagged their location and assigned alleles. Data were compared at 7 multi-locus sequence typing (MLST) loci using the summary report provided by the BIGSdb software, and at 1,643 loci defined by the annotation of the NCTC11168 genome (22) of *C. jejuni.*
Indirect comparison
Assembly statistics were compared across the first 309 clinical isolates from the North East and Oxfordshire sequenced at PHE in the current study and past Sanger Institute sequenced isolates from Oxfordshire, between 2011 and 2015 and available on PubMLST.

### Results and Discussion

Direct comparison
Seven locus MLST types derived from whole genome sequencing at PHE and the Sanger Institute were identical in all 22 instances (Table 42).
Analysis of data at 1,643 annotated loci identified 1,553 loci with information across all 22 PHE results and 24 Sanger Institute results. This analysis clearly distinguished among the thirteen patients, with repeated isolates or DNA preparations from each clustering clearly and substantially differentiated from all other patients (Figure 40), with large differences between isolates from

different patients compared to differences between the same isolate sequenced more than once at one site or the same isolate sequenced at each site for which identity exceeded 99% across 1,553 genetic loci.
Allelic variation between single sample DNA aliquots sequenced and assembled at the two centres was up to 12 loci (0.7%). Allelic variation between any two patients ranged between 988 and 1,494 loci apart from one exception, patients one and two who had highly similar isolates, differing at just 16 loci, but even here the two patients could be separated reliably with mean allelic variation between the patients (mean 12.1) almost double that observed within patient (mean 6.6). These two cases shared infection with ST-464, a recently emerged clonal complex with substantial clonal expansion showing limited genetic variation across isolates (81) and so are closely related in this sense although with no evidence that they are part of a specific outbreak.
Allelic variation among samples sequenced at different sites (1-12 loci; 0.07%-0.78%) was greater than that detected among same site sequenced samples (0-7 loci; 0.0%-0.46%). These differences are minor and do not impact on the applications in the FSA project. All are also within the range of variation previously reported for same patient samples WGS at the same site (75).

Table 42. Sequence type and clonal complex data obtained from WGS analysis performed by the Sanger Institute (a) and PHE (b).
(a)

| Patient | | BIGSdb ID | Isolate | ST | Clonal Complex |
|---|---|---|---|---|---|
| 1 | Replicate DNA Sample | 25590 | OXC88028_SV | 5136 | ST-464 complex |
| 1 | Replicate DNA Sample | 25647 | OXC88028R_SV | 5136 | ST-464 complex |
| 2 | Replicate DNA Sample | 28898 | OXC8397_SV | 5136 | ST-464 complex |
| 2 | Replicate DNA Sample | 29066 | OXC8397R_SV | 5136 | ST-464 complex |
| 3 | Replicate DNA Sample | 31028 | OXC8898_SV | 2122 | ST-353 complex |
| 3 | Replicate DNA Sample | 31005 | OXC8898R_SV | 2122 | ST-353 complex |
| 4 | same patient | 12901 | OXC6566_SV | 5 | ST-353 complex |
| 4 | same patient | 12910 | OXC6575_SV | 5 | ST-353 complex |
| 5 | Replicate DNA Sample | 28899 | OXC8543_SV | 991 | ST-692 complex |
| 5 | Replicate DNA Sample | 29069 | OXC8543R_SV | 991 | ST-692 complex |
| 6 | same patient | 16346 | OXC6595_SV | 573 | ST-573 complex |
| 6 | same patient | 16360 | OXC6610_SV | 573 | ST-573 complex |
| 7 | Replicate DNA Sample | 22709 | OXC7218_SV | 6543 | ST-828 complex |
| 7 | Replicate DNA Sample | 25550 | OXC7218_R_SV | 6543 | ST-828 complex |
| 8 | same patient | 16263 | OXC6471_SV | 827 | ST-828 complex |
| 8 | same patient | 16264 | OXC6472_SV | 827 | ST-828 complex |
| 9 | - | 12911 | OXC6576_SV | 855 | ST-828 complex |
| 10 | same patient | 16276 | OXC6484_SV | 403 | ST-403 complex |
| 10 | same patient | 16277 | OXC6485_SV | 403 | ST-403 complex |
| 11 | - | 12913 | OXC6578_SV | 312 | ST-658 complex |
| 12 | Replicate DNA Sample | 28982 | OXC8530_SV | 658 | ST-658 complex |
| 12 | Replicate DNA Sample | 29085 | OXC8530R_SV | 658 | ST-658 complex |
| 13 | Replicate DNA Sample | 28981 | OXC8405_SV | 50 | ST-21 complex |
| 13 | Replicate DNA Sample | 29067 | OXC8405R_SV | 50 | ST-21 complex |

(b)

| Patient | | BIGSdb ID | PHE ref no | Isolate | ST | Clonal Complex |
|---|---|---|---|---|---|---|
| 1 | Replicate DNA Sample | - | - | - | - | - |
| 1 | Replicate DNA Sample | 33275 | RL15000558 | OXC88028R_SV | 5136 | ST-464 complex |
| 2 | Replicate DNA Sample | 33276 | RL15000559 | OXC8397_SV | 5136 | ST-464 complex |
| 2 | Replicate DNA Sample | 33277 | RL15000560 | OXC8397R_SV | 5136 | ST-464 complex |
| 3 | Replicate DNA Sample | 33282 | RL15000567 | OXC8898_SV | 2122 | ST-353 complex |
| 3 | Replicate DNA Sample | 33314 | RL15000568 | OXC8898R_SV | 2122 | ST-353 complex |
| 4 | same patient | 33268 | RL15000549 | OXC6566_SV | 5 | ST-353 complex |
| 4 | same patient | 33269 | RL15000550 | OXC6575_SV | 5 | ST-353 complex |
| 5 | Replicate DNA Sample | 33280 | RL15000565 | OXC8543_SV | 991 | ST-692 complex |
| 5 | Replicate DNA Sample | 33281 | RL15000566 | OXC8543R_SV | 991 | ST-692 complex |
| 6 | same patient | 33272 | RL15000553 | OXC6595_SV | 573 | ST-573 complex |
| 6 | same patient | 33312 | RL15000554 | OXC6610_SV | 573 | ST-573 complex |
| 7 | Replicate DNA Sample | 33273 | RL15000555 | OXC7218_SV | 6543 | ST-828 complex |
| 7 | Replicate DNA Sample | 33274 | RL15000556 | OXC7218_R_SV | 6543 | ST-828 complex |
| 8 | same patient | 33264 | RL15000545 | OXC6471_SV | 827 | ST-828 complex |
| 8 | same patient | 33265 | RL15000546 | OXC6472_SV | 827 | ST-828 complex |
| 9 | - | 33270 | RL15000551 | OXC6576_SV | 855 | ST-828 complex |
| 10 | same patient | 33266 | RL15000547 | OXC6484_SV | 403 | ST-403 complex |
| 10 | same patient | 33267 | RL15000548 | OXC6485_SV | 403 | ST-403 complex |
| 11 | - | 33271 | RL15000552 | OXC6578_SV | 312 | ST-658 complex |
| 12 | Replicate DNA Sample | 33279 | RL15000563 | OXC8530_SV | 658 | ST-658 complex |
| 12 | Replicate DNA Sample | 33313 | RL15000564 | OXC8530R_SV | 658 | ST-658 complex |
| 13 | Replicate DNA Sample | - | - | - | - | - |
| 13 | Replicate DNA Sample | 33278 | RL15000562 | OXC8405R_SV | 50 | ST-21 complex |

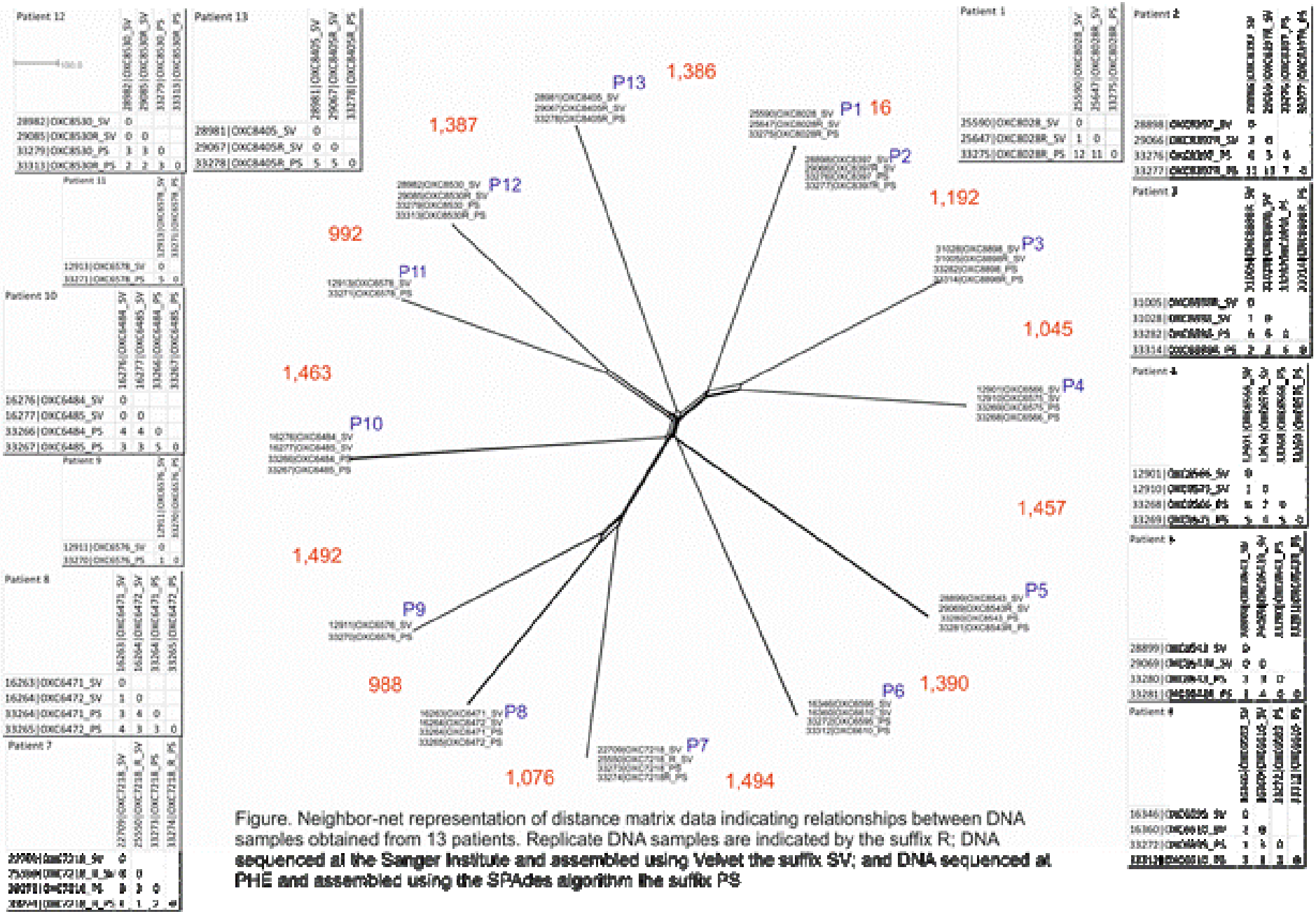



Figure. Neighbor-net representation of distance matrix data indicating relationships between DNA samples obtained from 13 patients. Replicate DNA samples are indicated by the suffix R; DNA **sequenced at the Sanger Institute and assembled using Velvet the suffix SV; and DNA sequenced at PHE and assembled using the SPAdes algorithm the suffix PS**

Figure 40. Phylogenetic tree illustrating the closeness between re-sequenced strains relative to isolates from other patients.

Indirect comparison
Graphic comparisons of assembly statistics are made in
Figure 41 and summarised in

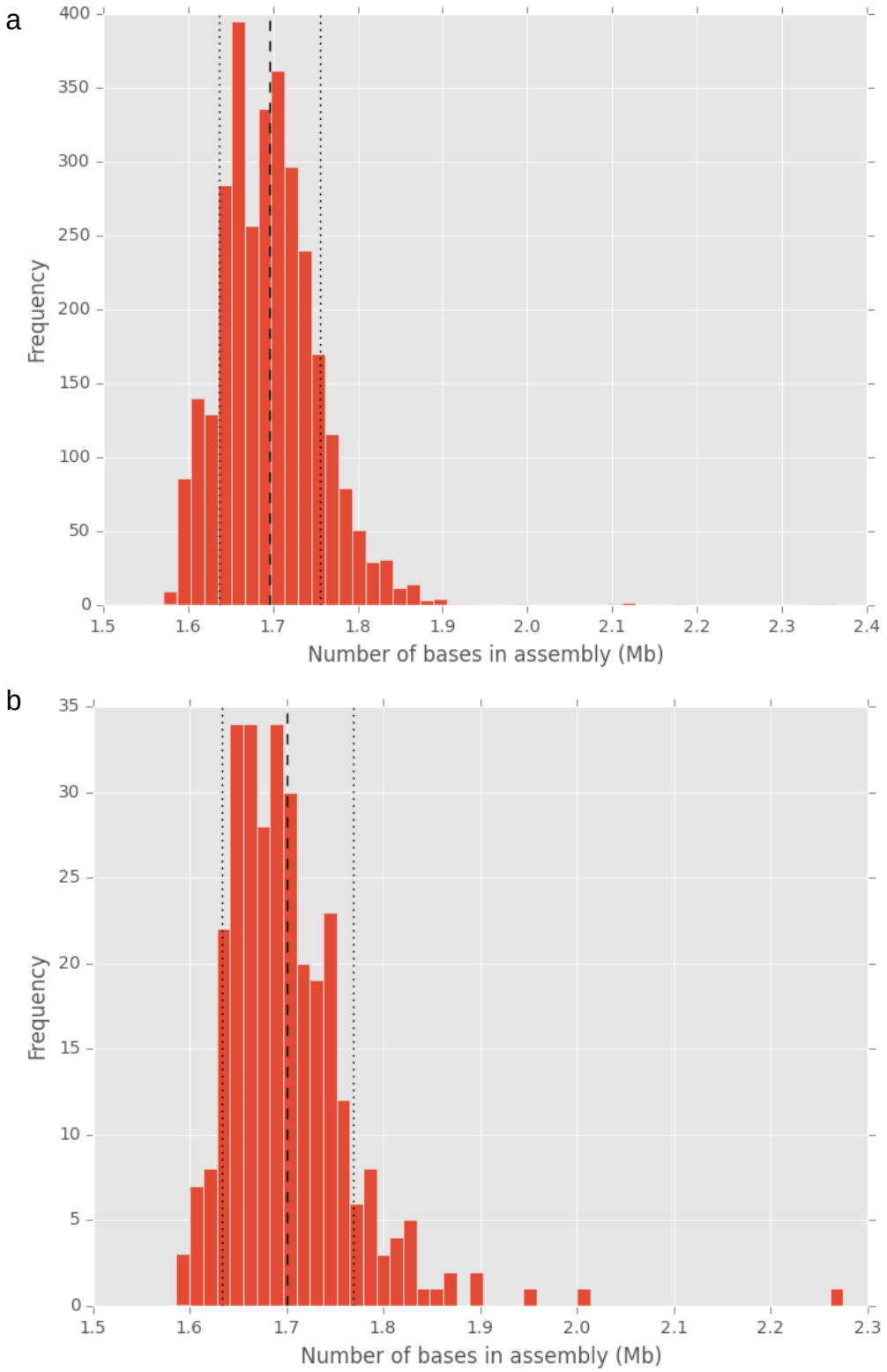
Table 43 Mean values are similar

## Summary

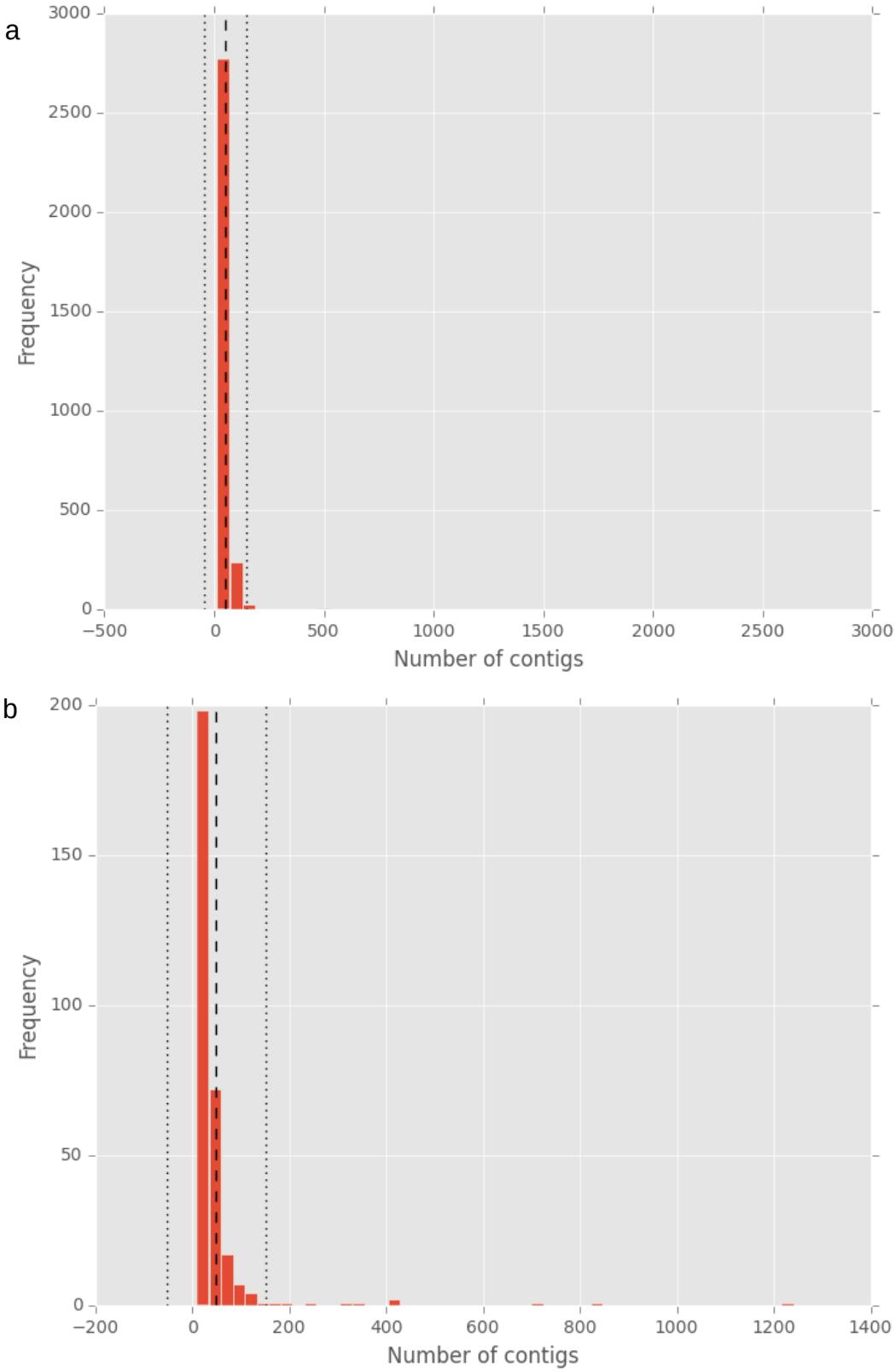
- WGS sequencing results from DNA aliquots performed at the Sanger Institute using Covaris library preparation and assembled with Velvet were highly similar to those sequenced at PHE following Nextera sample preparation and assembled using the SPAdes algorithm.
- Data from both centres reliably identified isolates at the MLST level of analysis.
- Sequences from DNA aliquots obtained from the two different centres showed identity at over 99% of 1,553 loci.
- This level of accuracy supported discrimination of isolates without artificially separating isolates from the same patient processed at different centres, which is the requirement to support outbreak detection. This level of accuracy will also exceed any expected accuracy and reproducibility requirement for whole genome sequence based source attribution.
- These results provided reassurance of the interchangeability of data from *Campylobacter* isolates processed by the Sanger Institute and PHE standard protocols and pipelines for all intended analyses.

**Assembly length**

Figure 41. Comparison of assembly statistics in the current study sequenced at PHE (a) and past work at the Sanger Institute (b).

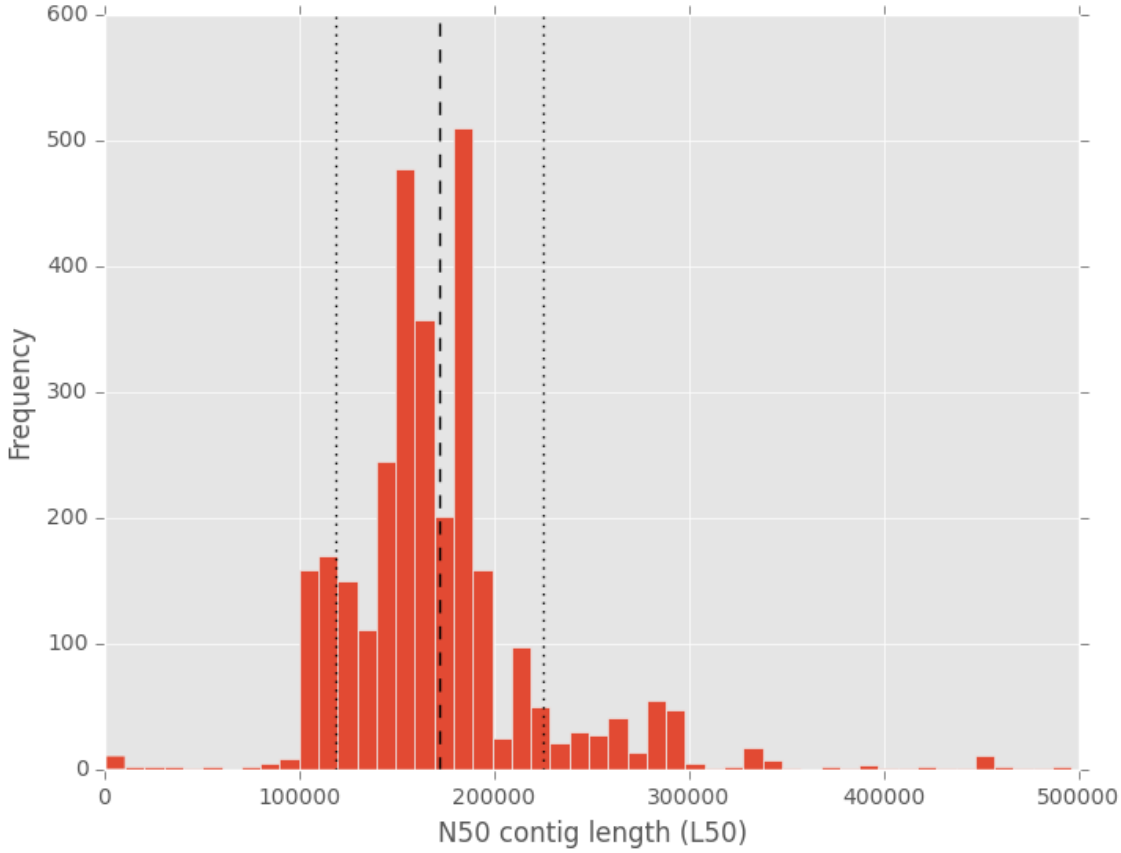


**Number of contigs**



**$N_{50}$ contig length ($L_{50}$)**

a




b

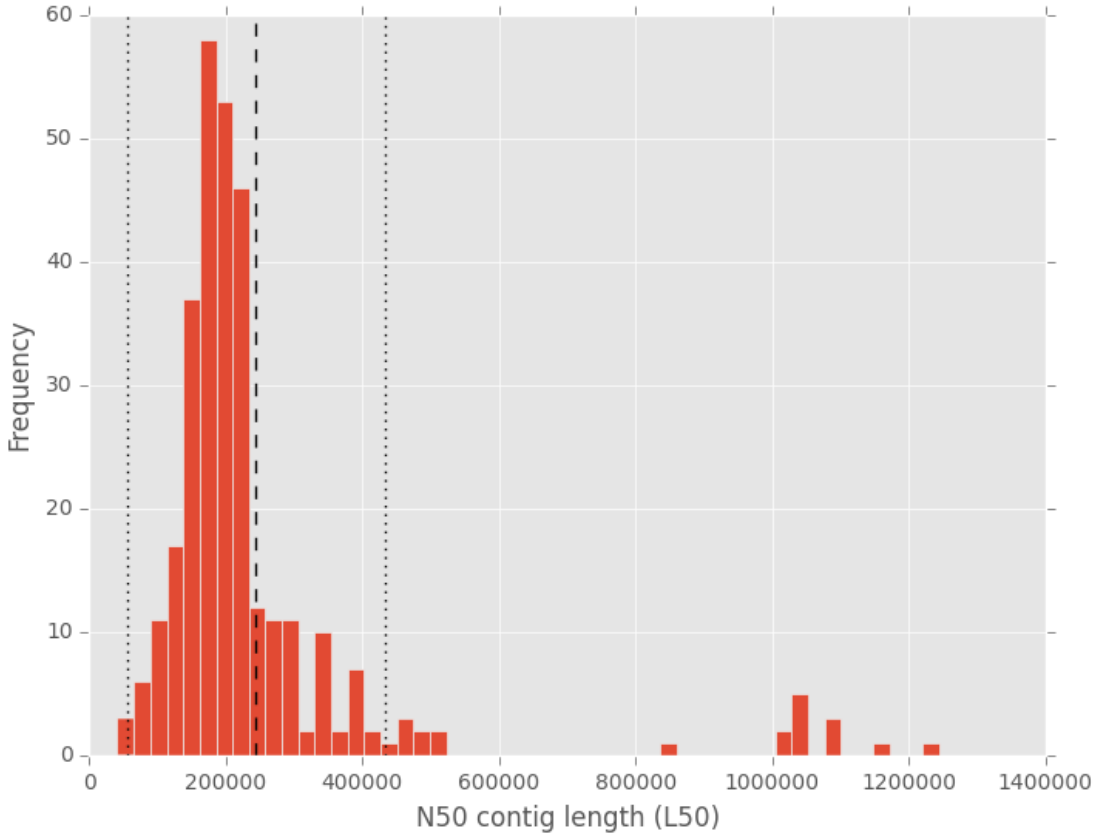


Table 43. Summary of assembly statistics and percent alleles designated / loci tagged based on the human disease *C. jejuni* / *C.coli* core genome MLST scheme v1.0.

FSA DATASET

| | $N_{50}$ length (bp) | Contigs (*n*) | Total length (Mb) | % alleles designated | % loci tagged |
|---|---|---|---|---|---|
| Mean | 243,892 | 50 | 1.70 | 96.2 | 96.2 |
| Std deviation | 188,415 | 102 | 0.07 | 4.6 | 4.6 |
| Minimum | 41,925 | 10 | 1.59 | 57.9 | 57.9 |
| 25% | 164,012 | 22 | 1.66 | 96.3 | 96.3 |
| 50% | 189,418 | 30 | 1.69 | 97.3 | 97.3 |
| 75% | 234,343 | 40 | 1.73 | 98.0 | 98 |
| Maximum | 1,244,251 | 1,241 | 2.28 | 99.3 | 99.3 |

OXFORDSHIRE 2011-2015

| | $N_{50}$ length (bp) | Contigs (*n*) | Total length (Mb) | % alleles designated | % loci tagged |
|---|---|---|---|---|---|
| Mean | 171,823 | 51 | 1.70 | 97.8 | 97.8 |
| Std deviation | 53,100 | 97 | 0.06 | 2.5 | 2.5 |
| Minimum | 696 | 14 | 1.57 | 20.9 | 20.9 |
| 25% | 147,075 | 33 | 1.66 | 97.5 | 97.5 |
| 50% | 162,660 | 41 | 1.69 | 97.8 | 97.8 |
| 75% | 185,751 | 53 | 1.73 | 98.4 | 98.4 |
| Maximum | 496,268 | 2,868 | 2.36 | 99.7 | 99.7 |

## 13.3 Appendix 3. Matching of notifications, questionnaires, and isolates for patient episodes of infection.

In total there were 6,160 records with a notification, a laboratory sample, or both. Unlinked samples mainly arose from individuals outside the study catchment areas having samples submitted to the laboratory. Notifications without a sample were a mix of notifications for patients in the study areas with laboratory samples submitted elsewhere and some samples not being forwarded from the clinical laboratories to Public Health England. Some individuals had several samples and some had multiple notifications of infection. Where samples or notifications were within one month of each other they were considered to be part of the same episode of infection.

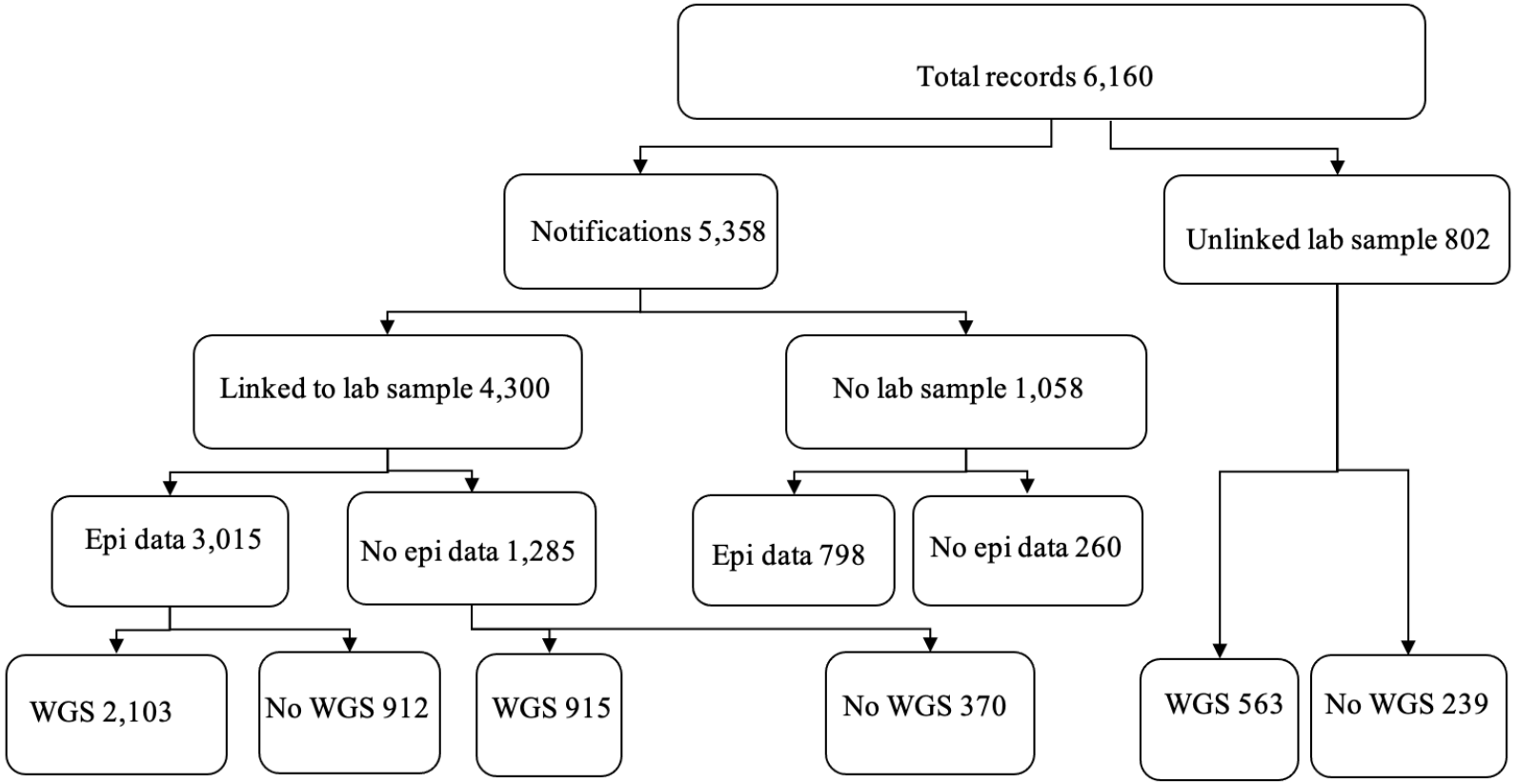



Figure 42. Relationship of notifications, laboratory samples, and epidemiological data in the study.

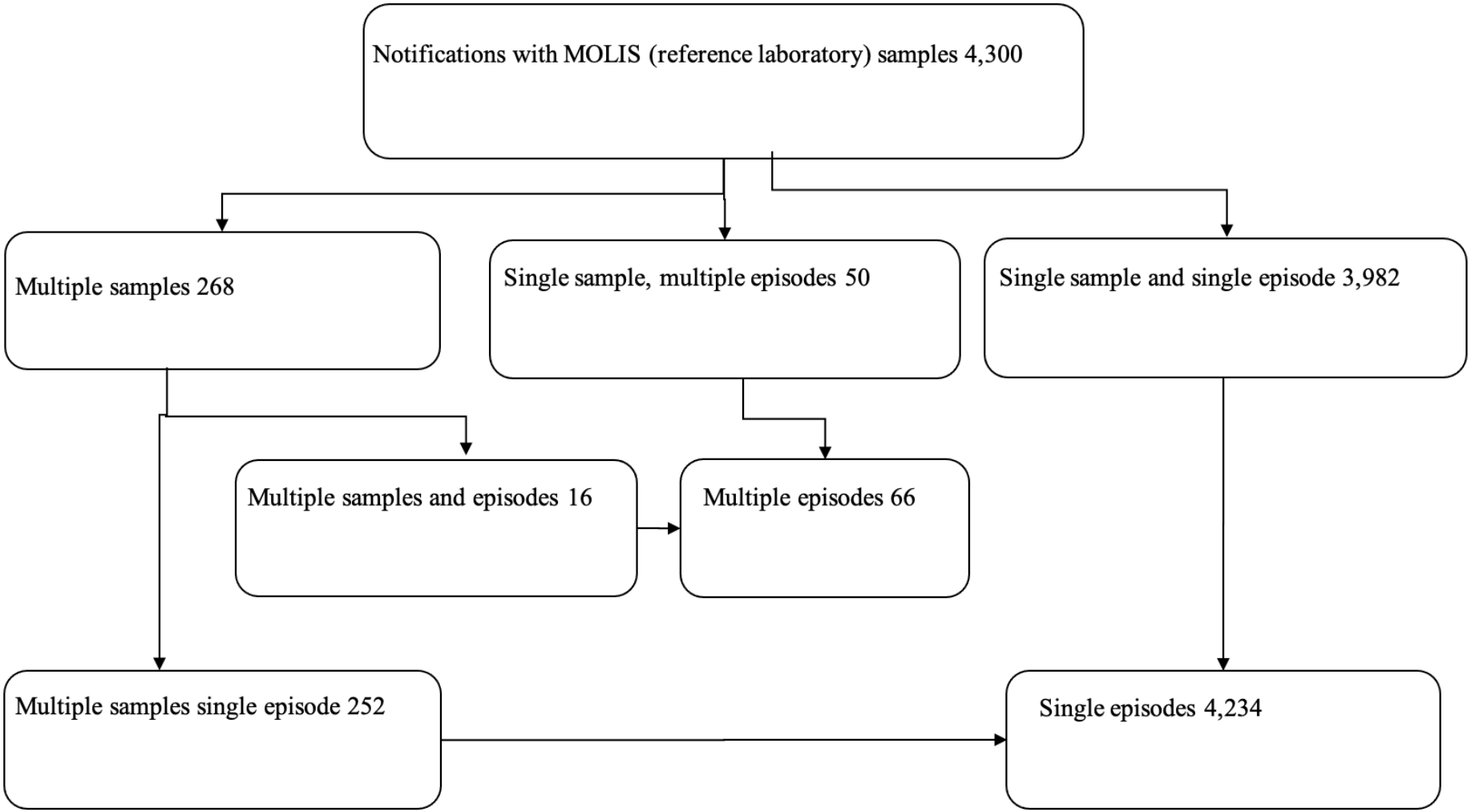



Figure 43. Linked notifications and laboratory samples identifying the extent and pattern of multiple samples and multiple notifications for individual patients

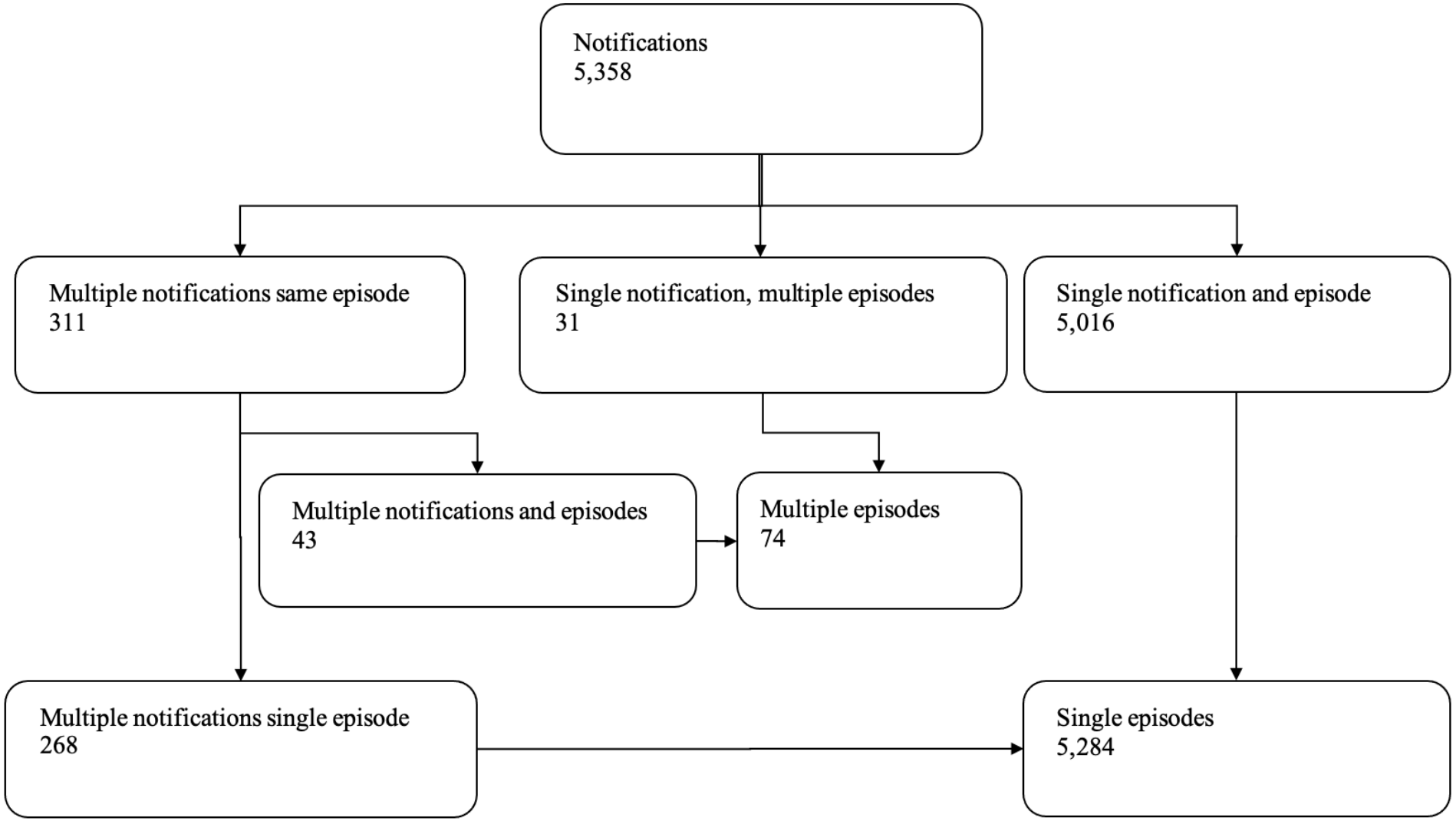



Figure 44. Relationship of notifications and episodes (treating two notifications within one month as a single episode).

## 13.4 Appendix 4. Additional tables and figures from descriptive epidemiology report.

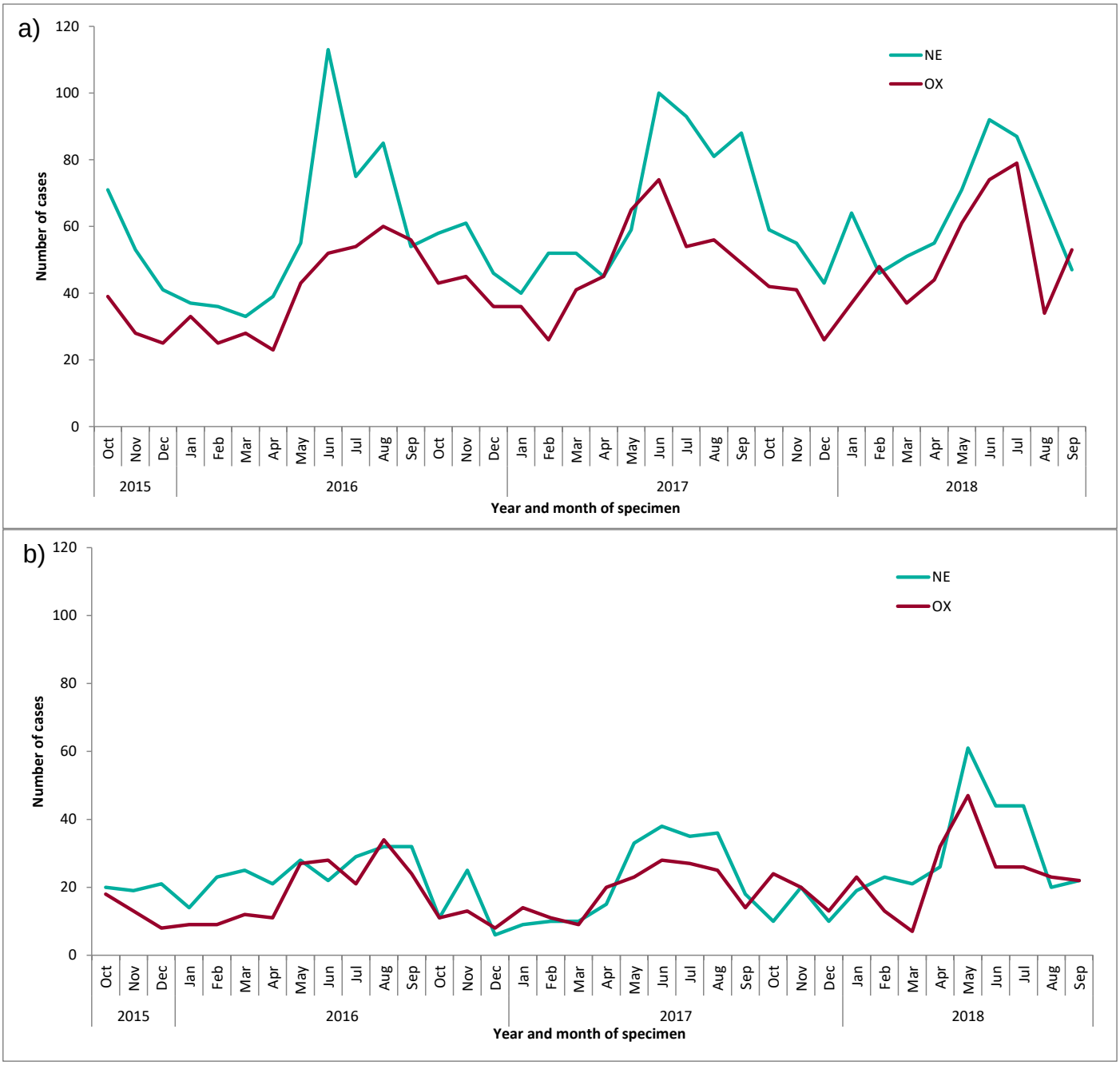



Figure 45: Case notification from specimen collection date of cases over the study period a) Respondents (n=3,816) b) Non-respondents (n=1,545)

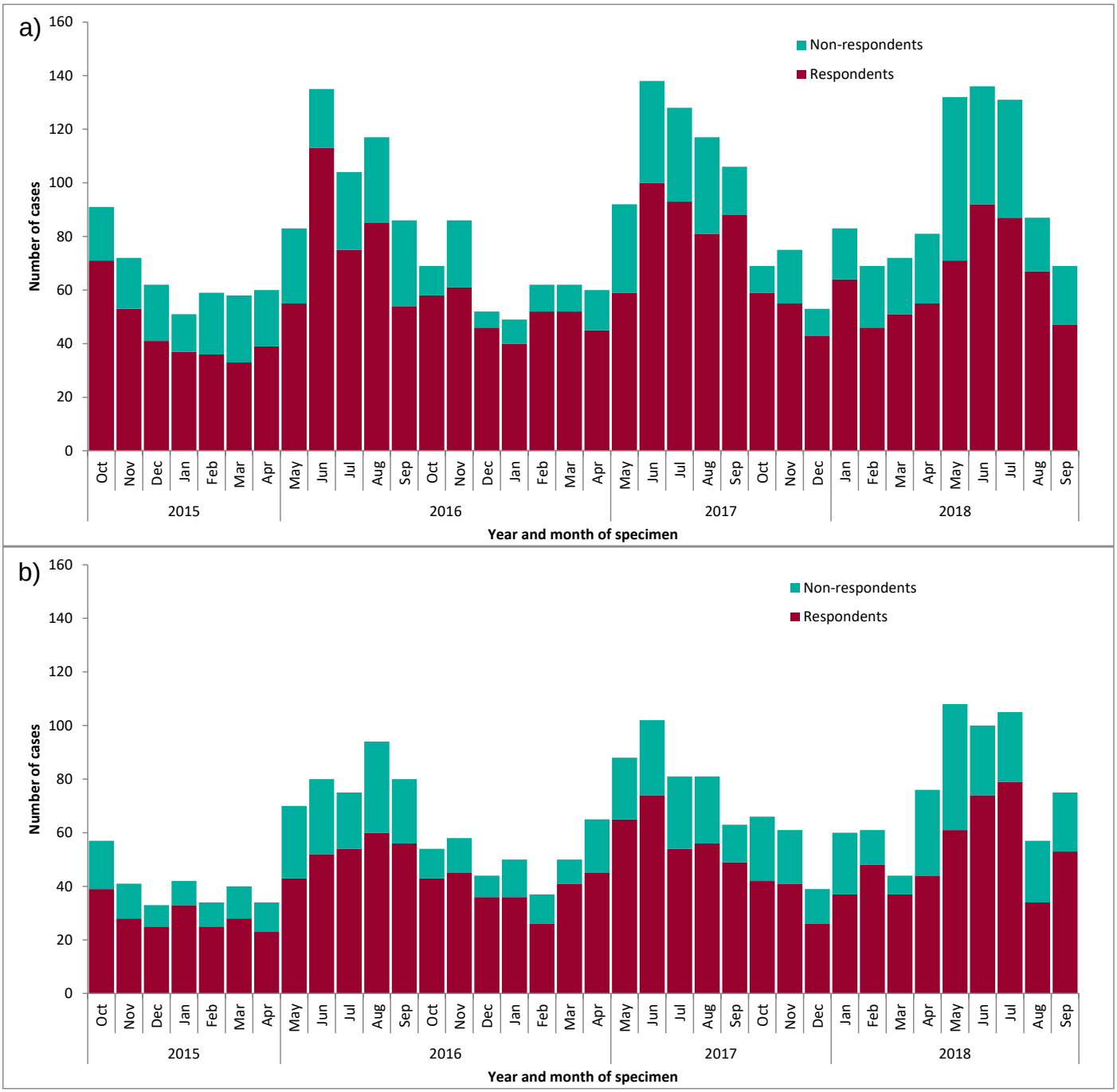
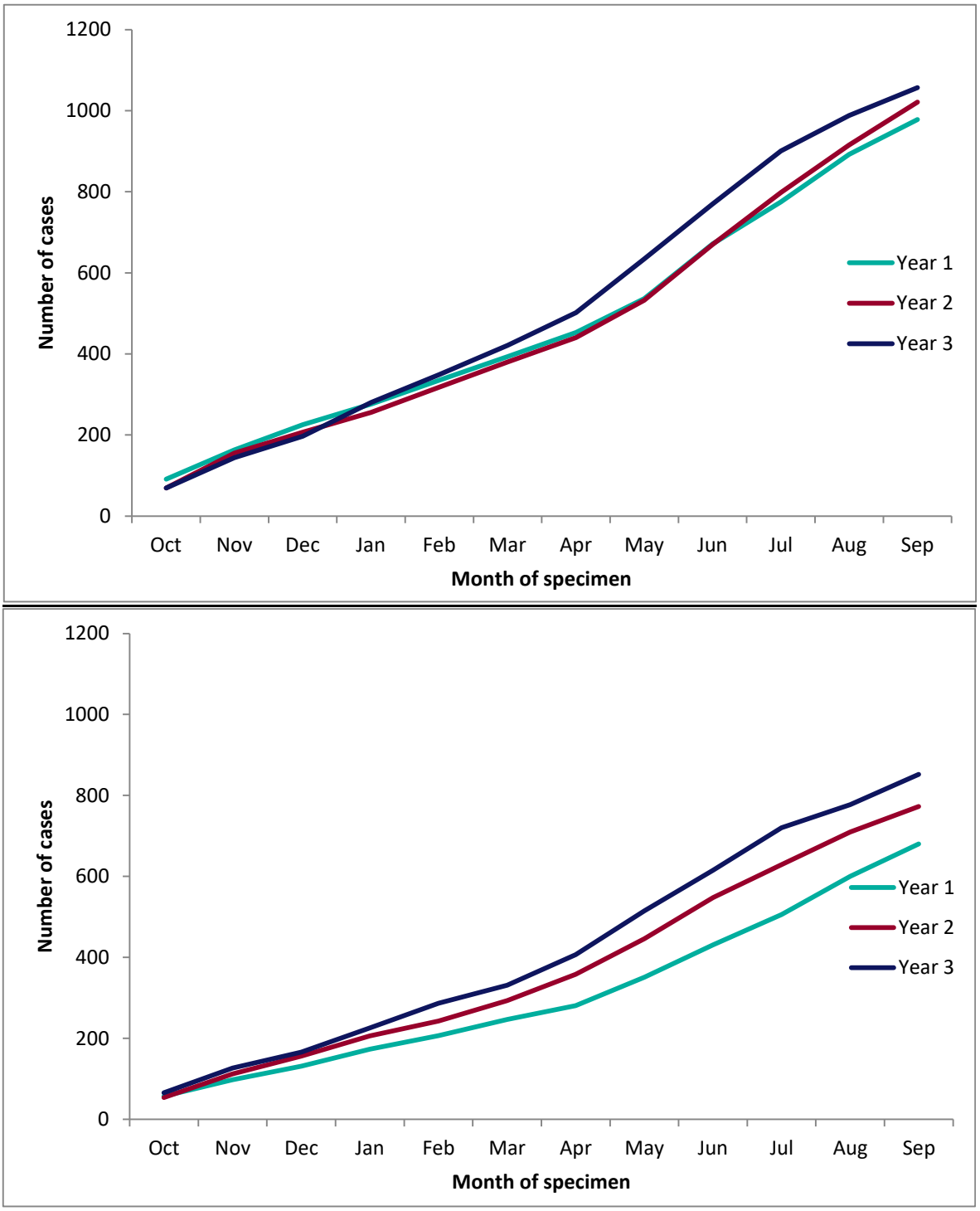



Figure 46: Case notification from specimen collection date of cases over the study period a) North East (n=3,056) b) Oxfordshire (n=2,305)

a)

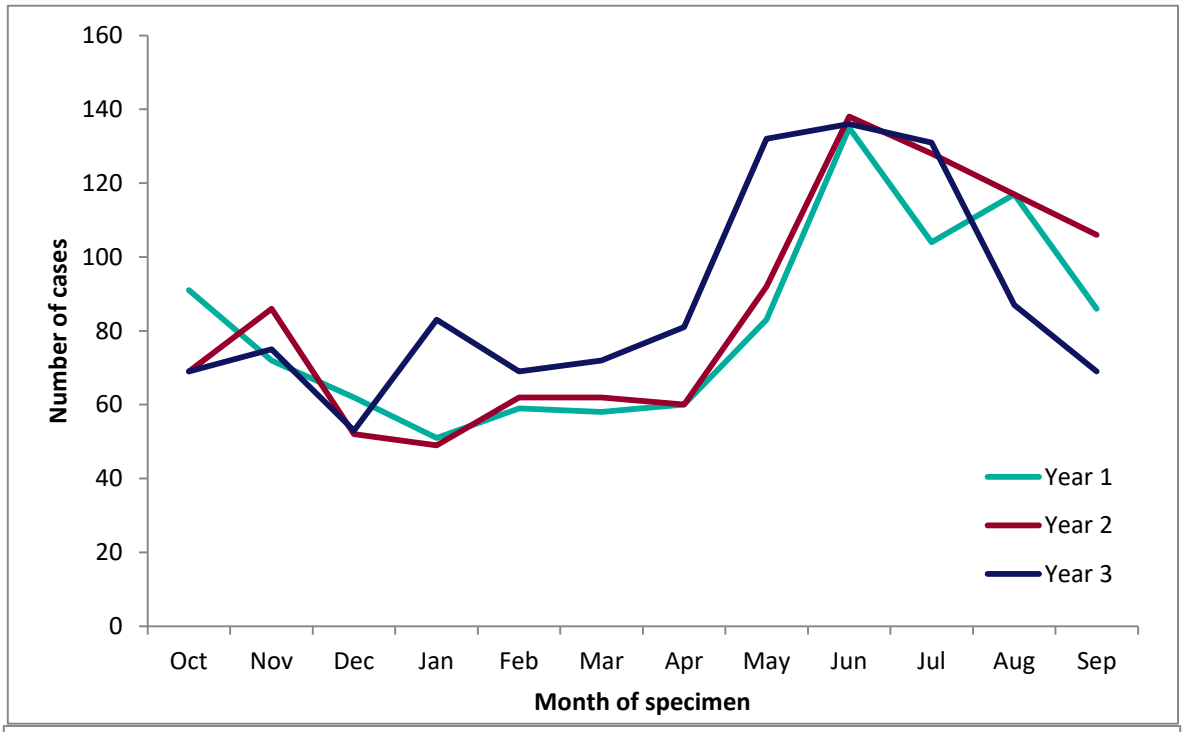



b)




Figure 47: Cumulative case notifications per month (from specimen collection date) by study year a) North East (n=3,056) b) Oxfordshire (n=2,305)

a)




b)

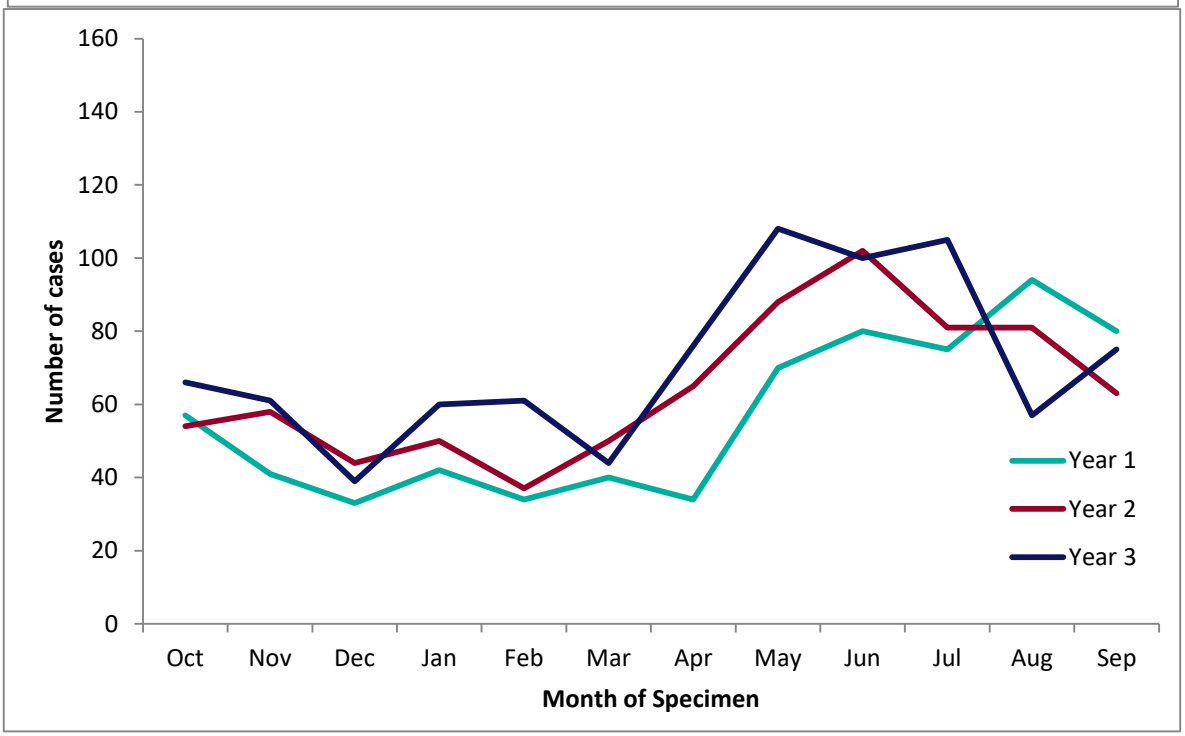



Figure 48. Number of case notifications per month (from specimen collection date) by study year a) North East (n=3,056) b) Oxfordshire (n=2,305).

Table 44: Break down of responses on food exposures for the whole dataset (n=3,821)

| Food | Yes-at home (n) | Yes-at home (%) | Yes-outside home (n) | Yes-outside home (%) | Yes-both (n) | Yes-both (%) | None (n) | None (%) | Total responses (n) | Total responses (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Chicken | 1416 | | 946 | 24.8 | 363 | 9.5 | 735 | 19.2 | 3460 | 90.6 |
| Duck, Turkey or Goose | 94 | 2.5 | 136 | 5.6 | 5 | 0.1 | 2668 | 69.8 | 2903 | 76.0 |
| Game birds | 12 | 0.3 | 14 | 0.6 | 0 | 0.0 | 2840 | 74.3 | 2866 | 75.0 |
| Lamb's liver | 85 | 2.2 | 35 | 1.5 | 3 | 0.1 | 2774 | 72.6 | 2897 | 75.8 |
| Cow or calf liver | 18 | 0.5 | 46 | 1.9 | 1 | 0.0 | 2807 | 73.5 | 2872 | 75.2 |
| Pig liver | 20 | 0.5 | 7 | 0.3 | 0 | 0.0 | 2833 | 74.1 | 2860 | 74.8 |
| Chicken liver | 30 | 0.8 | 39 | 1.6 | 1 | 0.0 | 2808 | 73.5 | 2878 | 75.3 |
| Liver pate or parfait | 73 | 1.9 | 61 | 2.5 | 4 | 0.1 | 2755 | 72.1 | 2893 | 75.7 |
| Unpasteurised/ raw milk | 17 | 0.4 | 47 | 1.9 | 3 | 0.1 | 2798 | 73.2 | 2865 | 75.0 |
| Any cold milk | 870 | 22.8 | 123 | 5.1 | 73 | 1.9 | 1992 | 52.1 | 3058 | 80.0 |

Table 45: Break down of responses on food exposures for the North East study site (n=2,206)

| Food | Yes-at home (n) | Yes-at home (%) | Yes-outside home (n) | Yes-outside home (%) | Yes-both (n) | Yes-both (%) | None (n) | None (%) | Total responses (n) | Total responses (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Chicken | 838 | 38.0 | 537 | 24.3 | 200 | 9.1 | 430 | 19.5 | 2005 | 90.9 |
| Duck, Turkey or Goose | 48 | 2.2 | 69 | 3.1 | 2 | 0.1 | 1548 | 70.2 | 1667 | 75.6 |
| Game birds | 6 | 0.3 | 4 | 0.2 | 0 | 0.0 | 1635 | 74.1 | 1645 | 74.6 |
| Lamb's liver | 60 | 2.7 | 19 | 0.9 | 2 | 0.1 | 1589 | 72.0 | 1670 | 75.7 |
| Cow or calf liver | 10 | 0.5 | 15 | 0.7 | 0 | 0.0 | 1626 | 73.7 | 1651 | 74.8 |
| Pig liver | 8 | 0.4 | 3 | 0.1 | 0 | 0.0 | 1635 | 74.1 | 1646 | 74.6 |
| Chicken liver | 15 | 0.7 | 21 | 1.0 | 1 | 0.0 | 1615 | 73.2 | 1652 | 74.9 |
| Liver pate or parfait | 36 | 1.6 | 26 | 1.2 | 1 | 0.0 | 1597 | 72.4 | 1660 | 75.2 |
| Unpasteurised/ raw milk | 12 | 0.5 | 23 | 1.0 | 1 | 0.0 | 1615 | 73.2 | 1651 | 74.8 |
| Any cold milk | 485 | 22.0 | 62 | 2.8 | 36 | 1.6 | 1176 | 53.3 | 1759 | 79.7 |

Table 46: Break down of responses on food exposures for the Oxfordshire study site (n=1,615)

| Food | Yes-at home (n) | Yes-at home (%) | Yes-outside home (n) | Yes-outside home (%) | Yes-both (n) | Yes-both (%) | None (n) | None (%) | Total responses (n) | Total responses (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Chicken | 578 | 35.8 | 409 | 25.3 | 163 | 10.1 | 305 | 18.9 | 1455 | 90.1 |
| Duck, Turkey or Goose | 46 | 2.8 | 67 | 4.1 | 3 | 0.2 | 1120 | 69.3 | 1236 | 76.5 |
| Game birds | 6 | 0.4 | 10 | 0.6 | 0 | 0.0 | 1205 | 74.6 | 1221 | 75.6 |
| Lamb's liver | 25 | 1.5 | 16 | 1.0 | 1 | 0.1 | 1185 | 73.4 | 1227 | 76.0 |
| Cow or calf liver | 8 | 0.5 | 31 | 1.9 | 1 | 0.1 | 1181 | 73.1 | 1221 | 75.6 |
| Pig liver | 12 | 0.7 | 4 | 0.2 | 0 | 0.0 | 1198 | 74.2 | 1214 | 75.2 |
| Chicken liver | 15 | 0.9 | 18 | 1.1 | 0 | 0.0 | 1193 | 73.9 | 1226 | 75.9 |
| Liver pate or parfait | 37 | 2.3 | 35 | 2.2 | 3 | 0.2 | 1158 | 71.7 | 1233 | 76.3 |
| Unpasteurised/ raw milk | 5 | 0.3 | 24 | 1.5 | 2 | 0.1 | 1183 | 73.3 | 1214 | 75.2 |
| Any cold milk | 385 | 23.8 | 61 | 3.8 | 37 | 2.3 | 816 | 50.5 | 1299 | 80.4 |

## 13.5 Appendix 5. Whole genome sequencing attribution – literature reviews

Completed 4 March 2016 and updated 1 February 2018

**Introduction**

This project aimed to evaluate the application of whole genome data to source attribution by applying any promising and relevant methods developed by others. To support this the population genetic host attribution literature was reviewed.

**Method**

PubMed searches using the text words: whole genome sequence attribution (12 results), genomic attribution infection (23), genome attribution infection (16 results), genomic attribution food (23), genome attribution food (17) were used to identify papers outside the study group's own work in this area. Titles were and abstracts were reviewed identifying eight papers of potential relevance.

Update February 2018

whole genome sequence attribution (12 + 11 new results, 1 remaining post title and abstract screening),

genomic attribution infection (23 + 7 new results, 4 remaining post title and abstract screening and deduplication),

genome attribution infection (16 results + 10 new results, 0 remaining post title and abstract screening and deduplication),

genomic attribution food (23+ 10 new results, 0 remaining post title and abstract screening and deduplication),

genome attribution food (17+ 10 new results, 0 remaining post title and abstract screening and deduplication)

**Results**

**4 March 2016**

No validated or population genetic method using whole genome data for source attribution was identified. Three papers recommended phylogenetic approaches to identification of source and including human to human transmission. One of these was a review based on expected signals from other studies (82), another a paper attributing to a specific local source (83) and none validated the method. One further review considered the area generally without a recommendation in the area (84) and one a database indicating forms of data that may be useful without a specific method (85).

Three papers (two from searches and one (78) from work by the study team) identified genes or sets of genes that may be of importance in host attribution beyond seven locus multi-locus

sequence type data (78, 86, 87). Two of these considered the impact of a single gene (86, 87) and the other a process to identify sets of genes linked to host-species.

One paper estimated the extent of switching of generalist lineages between host species as an index of the likely strength of host signal that genomes of these lineages are likely to carry. This study concluded that such switching is substantial and that reference datasets from putative sources may need to be extensive to allow accurate estimation of source.

**1 February 2018**

Five new papers describing approaches to using whole genomes to infer sources of isolates were identified. One searched the genome for genes with a strong geographical association and then validated those identified in an independent dataset (88). Seven genes were selected and provided 73% accuracy compared of estimating whether isolates were from domestic or travel associated cases (compared to 50% accuracy with no genetic information). Although this work shows some geographical structure in genetic make-up the method does not offer a real whole-genome based analytical approach. Another takes a similar approach to attribution to host, selecting those genes that perform best and starting with a highly conserved core of 472 genes (89). Self-attribution using these genes showed accuracy for attributing known chicken isolates to chicken of 73% to 77%, similar to 7-gene MLST based on large reference datasets, and with bias in favour of ruminant sources. Selection of gene groups based on criteria of lack of bias between ruminant and chicken produced lower overall accuracy and there is no validation on an independent dataset so that this approach does not currently offer an obvious advantage of 7-gene MLST based attribution based on available large datasets. Further development of this method may perform better when large whole genome datasets are available to allow independent datasets to support gene selection and validation. Two papers consider attribution of *Legionella* to source. One for *L. longbeachae* is based on a very small dataset and highlights limitations due to the high levels of diversity in reservoir populations and does not develop any specific method (90). The other for *L. pneumophila* used both phylogenetic and supervised statistical learning approaches to map cases to individual cooling towers and human cases from known outbreaks (91). Supervised statistical learning was applied to both SNP and core genome MLST data. Supervised statistical learning was reported to perform well with 93% accuracy when based on SNP data where identified recombination has been removed. The authors identified limitations due to not having a separate dataset to validate the method and report it as a test of principle. The phylogenetic data shows marked genetic structuring indicating that the task of attribution was supported by highly structured data. Finally, one paper applied machine learning approaches, but

based on gene presence and absence rather than sequence data in the context of application to pathogens which show a large amount or informative variation in gene content (92).

**Discussion**

No methods had been published before March 2016 for evaluation. Since then, of the new approaches published, the supervised statistical learning approach is the most likely candidate to trial for whole genome attribution. Other papers offer insight into which genes may perform best and the evidence for host switching identified in the March 2016 review should be considered in guiding the design and interpretation of attribution using genomes in these lineages but none of these is a candidate method.

We followed up with the corresponding author (Timothy Stinear) of the supervised statistical learning approach regarding application to *Campylobacter* source attribution in the context of our relatively large available datasets, but with potentially less identifiable source related population structure than the datasets on which the approach was piloted. He did not see this as an approach worth pursuing in collaboration and referred us to a related paper on genomic analysis of a *Yersinia* outbreak including colleagues interested in *Campylobacter* attribution as potentially of interest (93). Discussion with Prof Nigel French from that paper has led to collaboration on his team applying whole genome based asymmetric island analysis to the study data to pilot this approach that his team has been developing. This work is ongoing.

## 13.6 Appendix 6. Attribution reference population datasets

Finalised attribution datasets used in this study are publicly available on the pubmlst.org/campylobacter database and are summarised in Table 47 and

Table 48.
Table 47. MLST attribution datasets used in this study.

| **Project No.** | **Project** | **Project Description** | **No of isolates** |
|---|---|---|---|
| 66 | FS101013 *C. coli* attribution ref dataset | Reference *C. coli* isolates from poultry, ruminants and pigs for clinical source attribution. | 3,521 |
| 67 | FS101013 *C. jejuni* attribution ref dataset | Reference *C. jejuni* isolates from chicken, ruminants and wild birds for clinical source attribution. | 7,715 |
| 68 | FS101013 *C. coli* isolates for attribution | Clinical *C. coli* isolates from 2015-2018 for attribution to probable source | 399 |
| 69 | FS101013 *C. jejuni* isolates for attribution | Clinical *C. jejuni* isolates from 2015-2018 for attribution to probable source. | 3,885 |
| 70 | FS101013 OXC Long *C. coli* | Oxfordshire clinical *C. coli* isolates from 2003 - 2018, for attribution to probable source. | 925 |
| 71 | FS101013 OXC Long *C. jejuni* | Oxfordshire clinical *C. jejuni* isolates from 2003 - 2018, for attribution to probable source. | 8,768 |
| 75 | FS101013 OXC *C. coli* for attribution | Clinical *C. coli* isolates from Oxfordshire between 2015-2018 for attribution to probable source. | 156 |
| 76 | FS101013 OXC *C. jejuni* for attribution | Clinical *C. jejuni* isolates from Oxfordshire between 2015-2018 for attribution to probable source. | 1,590 |
| 77 | FS101013 NWC *C. coli* for attribution | Clinical *C. coli* isolates from North East England between 2015-2018 for attribution to probable source. | 243 |
| 78 | FS101013 NWC *C. jejuni* for attribution | Clinical *C. jejuni* isolates from North East England between 2015-2018 for attribution to probable source. | 2,295 |

Table 48. Reference datasets for development of whole genome attribution methodologies

| **Project No.** | **Project** | **Project Description** | **No of isolates** |
|---|---|---|---|
| 79 | FS101013 PubMLST WGS *C. coli* isolates | Genome sequenced *C. coli* isolates from animals and foods to develop a whole genome sequence reference database from important sources of human infection | 1,323 |
| 80 | FS101013 PubMLST WGS *C. jejuni* isolates | Genome sequenced *C. jejuni* isolates from animals and foods to develop a whole genome sequence reference database from important sources of human infection | 4,944 |

Isolates belonging to each dataset can be accessed by going to the front page of the PubMLST 'Isolate database', and under 'Projects' selecting 'Main Projects', or clicking here. The resulting output provides an interactive table of major projects, of which 10 relate to FS101013 MLST attribution, and two are WGS reference datasets for the development of WGS attribution. An isolate list for each project can be obtained by selecting the appropriate link in the right hand column of the table. Details of individual isolates are visible by selecting the interactive id link in the left hand column of the isolate table(s).

## 13.7 Appendix 7. Datasets from this study

### Accessible virtual isolate archive on PubMLST.

Finalised clinical, abattoir and food datasets used in this study are publicly available on the pubmlst.org/campylobacter database and are summarised in
Table 49. Isolates belonging to each dataset can be accessed by going to the front page of the PubMLST 'Isolate database', and under 'Projects' selecting 'Main Projects', or here. The resulting output provides an interactive table of major projects. An isolate list for each project can be obtained by selecting the appropriate link in the right hand column of the table.

Table 49. Clinical, food and abattoir isolate datasets from FS101013.

| Project No. | Project | Project Description | No of isolates |
|---|---|---|---|
| 9 | FS101013 NWC Human Surveillance Yr1 | Clinical isolates from North East England, obtained between October 2015 - September 2016. | 916 |
| 41 | FS101013 NWC Human Surveillance Yr1 dd | Clinical isolates from North East England, obtained between October 2015 - September 2016, from which patient replicate samples have been excluded. | 888 |
| 46 | FS101013 NWC Human Surveillance Yr2 | Clinical isolates from North East England, obtained between October 2016 - September 2017. | 804 |
| 47 | FS101013 NWC Human Surveillance Yr2 dd | Clinical isolates from North East England, obtained between October 2016 - September 2017, from which patient replicate samples have been excluded. | 778 |
| 53 | FS101013 NWC Human Surveillance Yr3 | Clinical isolates from North East England, obtained between October 2017 - September 2018. | 828 |
| 54 | FS101013 NWC Human Surveillance Yr3 dd | Clinical isolates from North East England, obtained between October 2017 - September 2018, from which patient replicate samples have been excluded. | 808 |
| 40 | FS101013 OXC Human Surveillance Yr1 | Clinical isolates from Oxfordshire, obtained between October 2015 - September 2016. | 460 |
| 42 | FS101013 OXC Human Surveillance Yr1 dd | Clinical isolates from Oxfordshire, obtained between October 2015 - September 2016, from which patient replicate samples have been excluded. | 440 |
| 48 | FS101013 OXC Human Surveillance Yr2 | Clinical isolates from Oxfordshire, obtained between October 2016 - September 2017. | 618 |

| Project No. | Project | Project Description | No of isolates |
|---|---|---|---|
| 49 | FS101013 OXC Human Surveillance Yr2 dd | Clinical isolates from Oxfordshire, obtained between October 2016 - September 2017, from which patient replicate samples have been excluded. | 586 |
| 55 | FS101013 OXC Human Surveillance Yr3 | Clinical isolates from Oxfordshire, obtained between October 2017 - September 2018. | 640 |
| 56 | FS101013 OXC Human Surveillance Yr3 dd | Clinical isolates from Oxfordshire, obtained between October 2017 - September 2018, from which patient replicate samples have been excluded. | 622 |
| 10 | FS101013 PHE historical isolates 1997-98 | Clinical isolates from UK between 1997-1998 | 479 |
| 72 | FS101013 OXC Human surveillance 2003-4 | Clinical isolates from Oxfordshire human surveillance between mid-September 2003 and mid-September 2004. | 447 |
| 73 | FS101013 OXC Human surveillance 2003-4 dd | Clinical isolates from Oxfordshire human surveillance between mid-September 2003 and mid-September 2004, from which patient replicate samples have been excluded. | 429 |
| 64 | FS101013 OXC Human surveillance 2006-07 | Clinical isolates from Oxfordshire human surveillance between mid-September 2006 and mid-September 2007. | 569 |
| 74 | FS101013 OXC Human surveillance 2006-07dd | Clinical isolates from Oxfordshire human surveillance between mid-September 2006 and mid-September 2007, from which patient replicate samples have been excluded. | 552 |
| 52 | FS101013 Retail Food Isolates | Isolates from retail chicken, duck, turkey, beef or lamb offal or meat. | 835 |
| 31 | FS101013 Chicken carcass isolates | Chicken carcass isolates collected at abattoir as part of FS241051. | 1,078 |
| 34 | FS101013 Pooled chicken caecal isolates | Pooled chicken caecal isolates collected at abattoir as part of FS241051. | 334 |
| 38 | FS101013 Pig abattoir survey | Pig abattoir isolates. | 78 |

## Summary of PHE processing of epidemiological and laboratory data

Cases

- 5938 cases based on NHS number throughout study
    - 799 lab samples only
    - 1055 unlinked to lab samples

Records

- Unique episode IDs: 5,358
- Unique molis ID: 5,324
- Unique questionnaire ID: 3,914
- Unique episode ID with molis linked: 4312
- Unique molis with episode ID linked:4,508
- Unique questionnaire ID with molis linked: 3,101
- Linked, deduplicated molis: 4,300

Molis (PHE laboratory system) records, linked and de-duplicated (4,300)

- *Campylobacter* species:
    - *C. jejuni*: 3,548 (82.5%)
    - *C. coli*: 384 (9.0%)
    - Other *Campylobacter* 17 (0.4%)
- Mixed: 117 (2.7%)
- Contaminated: 151 (3.5%)
- No growth: 74 (1.7%)
- Not *Campylobacter* species: 9 (0.2%)
- With SNP address: 3,018 (70.2%)

De-duplicated

- Notifications: 5,358
- Notifications linked to molis: 4,300
    - With epidemiology: 3,015
    - Without epidemiology: 1,285
- Notifications unlinked to molis: 1,058
    - With epidemiology: 798
    - Without epidemiology: 260
- Molis without notification Epi: 802

### North East site

De-duplicated

- Notifications: 3,257
- Notifications linked to molis: 2,817
    - With epidemiology: 2,028
    - Without epidemiology: 789
- Notifications unlinked to molis: 236
    - With epidemiology: 173

    - Without epidemiology: 63
- Molis without notification epidemiology: 204

Molis records, linked and de-duplicated (2,817)

- *Campylobacter* species:
    - *C. jejuni*: 2,341 (83.1%)
    - *C. coli*: 254 (9.0%)
    - *C. Others*: 16 (0.5%)
- Mixed: 83 (2.7%)
- Contaminated: 58 (2.0%)
- No growth: 62 (2.2%)
- Not *Campylobacter* species: 3 (0.1%)
- With SNP address: 2,016 (71.6%)

### Oxfordshire site

De-duplicated

- Notifications: 2,903
- Notifications linked to molis: 1,483
    - With epi: 987
    - Without: 496
- Notifications unlinked to molis: 822
    - With epi: 625
    - Without: 197
- Molis without notification Epi: 598

Molis records, linked and de-duplicated (1,192)

- *Campylobacter* species:
    - *C. jejuni*: 1,207 (81.4%)
    - *C. coli*: 130 (8.8%)
    - *C. Others*: 1 (0.1%)
- Mixed: 34 (2.3%)
- Contaminated: 93 (6.2%)
- No growth: 12 (0.8%)
- Not *Campylobacter* species: 6 (0.4%)
- With SNP address: 1,002 (84.1%)

## 13.8 Appendix 8. Population genetic attribution results in tables.

Table 50. Annual attribution of human cases of *Campylobacter jejuni* across chicken, ruminant, and wild bird sources using the Structure algorithm and study attribution dataset. Years run from October to September and are numbered as the later year, 2004 indicting October 2003 to September 2004. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Year | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Wild Bird | Adjusted Chicken | Adjusted Ruminant | Adjusted Wild Bird |
|---|---|---|---|---|---|---|---|
| 2004 | 558 | 58.3% | 39.6% | 2.1% | 75.2% | 24.8% | 0.0% |
| 2005 | 484 | 59.1% | 39.3% | 1.6% | 76.2% | 23.8% | 0.0% |
| 2006 | 483 | 59.9% | 38.4% | 1.7% | 77.5% | 22.5% | 0.0% |
| 2007 | 504 | 55.8% | 42.1% | 2.0% | 71.1% | 28.9% | 0.0% |
| 2008 | 485 | 56.1% | 42.0% | 1.9% | 71.4% | 28.6% | 0.0% |
| 2009 | 522 | 53.4% | 43.4% | 3.2% | 67.5% | 31.6% | 1.0% |
| 2010 | 592 | 58.8% | 39.2% | 2.0% | 76.0% | 24.0% | 0.0% |
| 2011 | 812 | 56.4% | 41.5% | 2.1% | 72.1% | 27.9% | 0.0% |
| 2012 | 813 | 51.8% | 46.0% | 2.2% | 64.6% | 35.4% | 0.0% |
| 2013 | 640 | 52.5% | 46.2% | 1.3% | 65.1% | 34.9% | 0.0% |
| 2014 | 715 | 52.4% | 45.9% | 1.7% | 65.2% | 34.8% | 0.0% |
| 2015 | 636 | 53.3% | 45.1% | 1.6% | 66.7% | 33.3% | 0.0% |
| 2016 | 445 | 53.1% | 45.5% | 1.3% | 66.2% | 33.8% | 0.0% |
| 2017 | 564 | 52.2% | 46.7% | 1.1% | 64.6% | 35.4% | 0.0% |
| 2018 | 579 | 56.4% | 42.3% | 1.3% | 71.6% | 28.4% | 0.0% |

Table 51. Annual attribution of human cases of *Campylobacter coli* across chicken, ruminant, and pig sources using the Structure algorithm and study attribution dataset. Years run from October to September and are numbered as the later year, 2004 indicting October 2003 to September 2004. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Year | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Pig | Adjusted Chicken | Adjusted Ruminant | Adjusted Pig |
|---|---|---|---|---|---|---|---|
| 2004 | 40 | 50.5% | 36.3% | 13.2% | 61.7% | 36.0% | 2.3% |
| 2005 | 50 | 52.3% | 36.8% | 10.9% | 64.1% | 35.9% | 0.0% |
| 2006 | 51 | 51.6% | 41.3% | 7.1% | 57.7% | 42.3% | 0.0% |
| 2007 | 61 | 48.8% | 34.5% | 16.6% | 59.8% | 32.8% | 7.4% |
| 2008 | 38 | 41.1% | 35.6% | 23.3% | 45.9% | 36.8% | 17.3% |
| 2009 | 31 | 49.1% | 37.2% | 13.7% | 58.9% | 38.2% | 2.9% |
| 2010 | 43 | 43.8% | 44.1% | 12.1% | 46.8% | 53.2% | 0.0% |
| 2011 | 105 | 49.3% | 37.9% | 12.8% | 59.0% | 39.4% | 1.5% |
| 2012 | 77 | 46.8% | 41.0% | 12.1% | 53.3% | 46.4% | 0.3% |
| 2013 | 107 | 45.4% | 44.5% | 10.0% | 48.0% | 52.0% | 0.0% |
| 2014 | 96 | 44.6% | 44.1% | 11.3% | 47.7% | 52.3% | 0.0% |
| 2015 | 71 | 42.3% | 43.6% | 14.1% | 44.6% | 52.4% | 3.0% |
| 2016 | 40 | 48.8% | 38.1% | 13.1% | 58.1% | 40.1% | 1.9% |
| 2017 | 53 | 44.4% | 47.2% | 8.4% | 43.8% | 56.2% | 0.0% |
| 2018 | 61 | 36.7% | 49.1% | 14.3% | 32.7% | 64.4% | 2.9% |

Table 52. Quarterly attribution of human cases of *Campylobacter jejuni* across chicken, ruminant, and wild bird sources using the Structure algorithm and study MLST attribution dataset. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Quarter | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Pig | Adjusted Chicken | Adjusted Ruminant | Adjusted Pig |
|---|---|---|---|---|---|---|---|
| 2015_4 | 34 | 45.8% | 42.1% | 12.2% | 51.0% | 48.7% | 0.3% |
| 2016_1 | 21 | 46.4% | 39.5% | 14.1% | 53.2% | 43.5% | 3.3% |
| 2016_2 | 32 | 29.7% | 57.1% | 13.2% | 18.0% | 81.1% | 0.8% |
| 2016_3 | 38 | 40.8% | 45.8% | 13.5% | 41.0% | 57.1% | 1.9% |
| 2016_4 | 30 | 45.1% | 45.7% | 9.3% | 46.3% | 53.7% | 0.0% |
| 2017_1 | 18 | 34.8% | 59.4% | 5.8% | 21.0% | 79.0% | 0.0% |
| 2017_2 | 37 | 33.4% | 53.2% | 13.4% | 25.6% | 73.0% | 1.4% |
| 2017_3 | 47 | 45.5% | 43.5% | 11.0% | 49.3% | 50.7% | 0.0% |
| 2017_4 | 37 | 30.1% | 54.1% | 15.7% | 19.8% | 75.6% | 4.6% |
| 2018_1 | 17 | 36.7% | 43.6% | 19.7% | 34.9% | 53.8% | 11.3% |
| 2018_2 | 40 | 30.9% | 49.2% | 19.9% | 22.8% | 66.2% | 11.0% |
| 2018_3 | 46 | 38.6% | 50.6% | 10.8% | 34.3% | 65.7% | 0.0% |

Table 53. Quarterly attribution of human cases of *Campylobacter coli* across chicken, ruminant, and pig sources using the Structure algorithm and study MLST attribution dataset. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Quarter | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Wild bird | Adjusted Chicken | Adjusted Ruminant | Adjusted Wild bird |
|---|---|---|---|---|---|---|---|
| 2015_4 | 312 | 47.6% | 51.2% | 1.2% | 57.0% | 43.0% | 0.0% |
| 2016_1 | 247 | 50.8% | 48.3% | 0.9% | 62.1% | 37.9% | 0.0% |
| 2016_2 | 335 | 50.6% | 47.2% | 2.2% | 62.6% | 37.4% | 0.0% |
| 2016_3 | 388 | 54.2% | 44.1% | 1.7% | 68.2% | 31.8% | 0.0% |
| 2016_4 | 315 | 48.5% | 50.4% | 1.2% | 58.4% | 41.6% | 0.0% |
| 2017_1 | 233 | 49.9% | 48.9% | 1.2% | 60.8% | 39.2% | 0.0% |
| 2017_2 | 385 | 55.1% | 42.2% | 2.7% | 70.2% | 29.5% | 0.3% |
| 2017_3 | 358 | 46.1% | 52.6% | 1.3% | 54.6% | 45.4% | 0.0% |
| 2017_4 | 309 | 45.7% | 52.7% | 1.5% | 54.2% | 45.8% | 0.0% |
| 2018_1 | 203 | 54.1% | 43.7% | 2.2% | 68.3% | 31.7% | 0.0% |
| 2018_2 | 449 | 59.6% | 38.5% | 1.9% | 77.3% | 22.7% | 0.0% |
| 2018_3 | 341 | 51.0% | 47.1% | 2.0% | 63.1% | 36.9% | 0.0% |

Table 54. Attribution by sentinel site and combined of human *Campylobacter jejuni* isolates from FS101013 to chicken, ruminant, and wild bird sources using the Structure algorithm and study MLST attribution dataset. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Sentinel site | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Wild bird | Adjusted Chicken | Adjusted Ruminant | Adjusted Wild bird |
|---|---|---|---|---|---|---|---|
| North East | 2,293 | 49.9% | 48.2% | 2.0% | 61.3% | 38.7% | 0.0% |
| Oxfordshire | 1,588 | 53.8% | 45.0% | 1.3% | 67.2% | 32.8% | 0.0% |
| Combined | 3,881 | 51.5% | 46.9% | 1.7% | 63.7% | 36.3% | 0.0% |

Table 55. Attribution by sentinel site and combined of human Campylobacter coli isolates from FS101013 to chicken, ruminant, and pig sources using the Structure algorithm and study MLST attribution dataset. Values are given with and without adjustment for bias estimated from validation using self-attribution.

| Sentinel site | N | Unadjusted Chicken | Unadjusted Ruminant | Unadjusted Pig | Adjusted Chicken | Adjusted Ruminant | Adjusted Pig |
|---|---|---|---|---|---|---|---|
| North East | 243 | 37.4% | 48.3% | 14.4% | 34.2% | 62.7% | 3.1% |
| Oxfordshire | 154 | 39.3% | 49.3% | 11.4% | 36.4% | 63.6% | 0.0% |
| Combined | 397 | 38.1% | 48.7% | 13.2% | 35.3% | 63.3% | 1.4% |

Table 56. Sequence types acting as sentinel types for chicken, cattle or ruminant attribution (CCna where a sequence type is not part of an identified clonal complex).

| CC and ST | Cattle | Sheep | Chicken | Wild bird | Total | Sentinel for: |
|---|---|---|---|---|---|---|
| CC443 ST 51 | 1 | 0 | 298 | 0 | 299 | Chicken |
| CC61 ST 61 | 188 | 66 | 20 | 0 | 274 | Ruminant |
| CC573 ST 573 | 0 | 0 | 232 | 0 | 232 | Chicken |
| CC661 ST 958 | 0 | 0 | 219 | 0 | 219 | Chicken |
| CC661 ST 814 | 0 | 0 | 173 | 0 | 173 | Chicken |
| CC574 ST 574 | 0 | 0 | 169 | 1 | 170 | Chicken |
| CC607 ST 607 | 0 | 0 | 133 | 0 | 133 | Chicken |
| CC354 ST 1489 | 0 | 0 | 108 | 0 | 108 | Chicken |
| CC354 ST 354 | 1 | 1 | 88 | 1 | 91 | Chicken |
| CC464 ST 464 | 2 | 0 | 76 | 1 | 79 | Chicken |
| CC607 ST 1212 | 1 | 0 | 77 | 0 | 78 | Chicken |
| CC283 ST 267 | 3 | 2 | 65 | 2 | 72 | Chicken |
| CC42 ST 459 | 68 | 0 | 2 | 0 | 70 | Cattle & Ruminant |
| CC48 ST 38 | 61 | 2 | 0 | 1 | 64 | Cattle & Ruminant |
| CC49 ST 49 | 4 | 0 | 57 | 0 | 61 | Chicken |
| CCna ST 1399 | 0 | 0 | 55 | 0 | 55 | Chicken |
| CC206 ST 572 | 4 | 0 | 49 | 0 | 53 | Chicken |
| CCna ST 2274 | 0 | 0 | 51 | 0 | 51 | Chicken |
| CC206 ST 206 | 19 | 26 | 3 | 0 | 48 | Ruminant |
| CC206 ST 122 | 4 | 0 | 40 | 0 | 44 | Chicken |
| CC353 ST 353 | 0 | 0 | 40 | 1 | 41 | Chicken |
| CC257 ST 2030 | 0 | 0 | 38 | 0 | 38 | Chicken |
| CC48 ST 475 | 1 | 0 | 36 | 0 | 37 | Chicken |
| CC692 ST 3120 | 0 | 0 | 36 | 0 | 36 | Chicken |
| CCna ST 1257 | 0 | 0 | 32 | 1 | 33 | Chicken |
| CC353 ST 5 | 1 | 0 | 28 | 0 | 29 | Chicken |
| CC45 ST 25 | 0 | 0 | 29 | 0 | 29 | Chicken |
| CC661 ST 1496 | 0 | 0 | 29 | 0 | 29 | Chicken |
| CCna ST 1492 | 0 | 0 | 29 | 0 | 29 | Chicken |
| CCna ST 1495 | 0 | 0 | 28 | 0 | 28 | Chicken |
| CC21 ST 982 | 26 | 0 | 1 | 0 | 27 | Cattle & Ruminant |
| CC45 ST 230 | 2 | 0 | 25 | 0 | 27 | Chicken |
| CC353 ST 5205 | 0 | 0 | 26 | 0 | 26 | Chicken |
| CCna ST 2217 | 25 | 0 | 0 | 0 | 25 | Cattle & Ruminant |
| CC21 ST 44 | 0 | 0 | 25 | 0 | 25 | Chicken |
| CC45 ST 538 | 0 | 0 | 24 | 1 | 25 | Chicken |

| CC and ST | Cattle | Sheep | Chicken | Wild bird | Total | Sentinel for: |
|---|---|---|---|---|---|---|
| CC443 ST 51 | 1 | 0 | 298 | 0 | 299 | Chicken |
| CC61 ST 61 | 188 | 66 | 20 | 0 | 274 | Ruminant |
| CC573 ST 573 | 0 | 0 | 232 | 0 | 232 | Chicken |
| CCna ST 58 | 23 | 0 | 0 | 0 | 23 | Cattle & Ruminant |
| CC1287 ST 945 | 0 | 0 | 23 | 0 | 23 | Chicken |
| CC257 ST 824 | 0 | 0 | 23 | 0 | 23 | Chicken |
| CC206 ST 46 | 0 | 0 | 22 | 0 | 22 | Chicken |
| CCna ST 6411 | 0 | 0 | 22 | 0 | 22 | Chicken |
| CC21 ST 883 | 2 | 0 | 19 | 0 | 21 | Chicken |
| CC403 ST 270 | 18 | 3 | 0 | 0 | 21 | Ruminant |
| CC607 ST 3963 | 0 | 0 | 21 | 0 | 21 | Chicken |
| CC464 ST 5136 | 0 | 0 | 20 | 0 | 20 | Chicken |
| CC354 ST 2863 | 0 | 0 | 19 | 0 | 19 | Chicken |
| CC354 ST 1073 | 1 | 0 | 17 | 0 | 18 | Chicken |
| CC21 ST 806 | 17 | 0 | 0 | 0 | 17 | Cattle & Ruminant |
| CC21 ST 148 | 0 | 0 | 17 | 0 | 17 | Chicken |
| CC353 ST 1232 | 0 | 0 | 17 | 0 | 17 | Chicken |
| CC61 ST 1244 | 16 | 0 | 0 | 0 | 16 | Cattle & Ruminant |
| CC354 ST 878 | 0 | 0 | 16 | 0 | 16 | Chicken |
| CC403 ST 55 | 15 | 0 | 0 | 0 | 15 | Cattle & Ruminant |
| CC257 ST 2254 | 0 | 0 | 15 | 0 | 15 | Chicken |
| CC283 ST 383 | 0 | 0 | 15 | 0 | 15 | Chicken |
| CC45 ST 233 | 1 | 0 | 14 | 0 | 15 | Chicken |
| CCna ST 905 | 1 | 0 | 14 | 0 | 15 | Chicken |
| CC403 ST 933 | 14 | 0 | 0 | 0 | 14 | Cattle & Ruminant |
| CC257 ST 584 | 0 | 0 | 14 | 0 | 14 | Chicken |
| CC353 ST 1036 | 1 | 0 | 13 | 0 | 14 | Chicken |
| CCna ST 877 | 0 | 0 | 14 | 0 | 14 | Chicken |
| CC353 ST 400 | 0 | 0 | 13 | 0 | 13 | Chicken |
| CC45 ST 2219 | 0 | 0 | 13 | 0 | 13 | Chicken |
| CC21 ST 2135 | 0 | 0 | 12 | 0 | 12 | Chicken |
| CC661 ST 661 | 0 | 0 | 12 | 0 | 12 | Chicken |
| CC42 ST 1013 | 11 | 0 | 0 | 0 | 11 | Cattle & Ruminant |
| CC353 ST 1210 | 0 | 0 | 11 | 0 | 11 | Chicken |
| CC460 ST 460 | 0 | 0 | 11 | 0 | 11 | Chicken |

| CC and ST | Cattle | Sheep | Chicken | Wild bird | Total | Sentinel for: |
|---|---|---|---|---|---|---|
| CC443 ST 51 | 1 | 0 | 298 | 0 | 299 | Chicken |
| CC61 ST 61 | 188 | 66 | 20 | 0 | 274 | Ruminant |
| CC573 ST 573 | 0 | 0 | 232 | 0 | 232 | Chicken |
| CC658 ST 1044 | 0 | 0 | 11 | 0 | 11 | Chicken |
| CCna ST 881 | 1 | 0 | 10 | 0 | 11 | Chicken |
| Total | 532 | 100 | 2,769 | 9 | 3,410 | - |

## 13.9 Appendix 9. Antimicrobial resistance estimated from genome data

### Comparison of BIGSdb results with PHE pipeline and access to individual results

Resistance prediction was identical by both approaches for macrolides and very similar for fluoroquinolones and tetracyclines with a maximum discordance of 0.4% for each. Table 57 show the percentages predicted as resistant by species and antimicrobial class.

Table 57. The percentage of fluoroquinolone, tetracycline and macrolide resistant clinical isolates (2015-2018) as detected by the PHE pipeline or BIGsDB software.

| Method | Fluoroquinolone PHE | Fluoroquinolone BIGSdb | Tetracycline PHE | Tetracycline BIGSdb | Macrolide PHE | Macrolide BIGSdb |
|---|---|---|---|---|---|---|
| *C. coli* | 38.1 | 37.9 | 36.6 | 36.6 | 4.1 | 4.1 |
| *C. jejuni* | 44.9 | 45.3 | 42.8 | 43.2 | 0.43 | 0.43 |

The predicted resistance for these three antimicrobial classes by each method is accessible via PubMLST for isolates from this FS101013 study. The naming convention is that macrolide_genotypes_1 gives the predicted resistance for macrolides from PHE, while substituting _2 for _1 gives BIGSdb software result. Similarly substituting “tetracycline” and “fluoroquinolone” for “macrolide” will identify the results for these antibiotic classes. This information can be viewed on each isolate record with an AMR conferring mutation or gene by selecting the appropriate ‘Phenotypic field’ from the drop down menu available on the ‘Search database’ page of the PubMLST database. “Phenotypic field” is a standard field on PubMLST for assignment of any characteristic to a record, which supports addition of phenotype. In this instance it is used as an available site for data and does not record phenotypic measures of resistance, but resistance inferred from the genome.

## Tabulation of data presented as figures in the main report

Table 58. Antimicrobial resistance among *Campylobacter jejuni* from human cases across four time periods in England and Wales.

| | England and Wales 1997/8 (n=461) | Oxfordshire 2003/4 (n=392) | Oxfordshire 2006/7 (n=494) | Sentinel study 2015-2018 (n=3,945) |
|---|---|---|---|---|
| Aminoglycoside | 5 (1%) | 4 (1%) | 4 (1%) | 27 (1%) |
| Fluoroquinolone | 24 (5%) | 101 (26%) | 158 (32%) | 1,780 (45%) |
| Macrolide | 0 (0%) | 0 (0%) | 2 (0%) | 17 (0%) |
| Tetracycline | 97 (21%) | 133 (34%) | 160 (32%) | 1,689 (43%) |

Table 59. Antimicrobial resistant among *Campylobacter jejuni* from food animal isolates sequenced as part of FS101013.

| | Chicken (abattoir) (n=889) | Chicken (retail) (n=401) | Ox/calf liver (n=99) | Lamb liver (n=137) | Duck (n=66) | Pigs (n=10) |
|---|---|---|---|---|---|---|
| Aminoglycoside | 0 (0%) | 1 (0%) | 2 (2%) | 0 (0%) | 0 (0%) | 1 (30%) |
| Fluoroquinolone | 380 (43%) | 210 (52%) | 13 (13%) | 11 (8%) | 6 (9%) | 0 (0%) |
| Macrolide | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Tetracycline | 499 (56%) | 243 (61%) | 14 (14%) | 9 (7%) | 39 (59% | 7 (70%) |

Table 60. Antimicrobial resistance among *Campylobacter coli* from human cases across four time periods in England and Wales.

| | England and Wales 1997/8 (n=18) | Oxfordshire 2003/4 (n=37) | Oxfordshire 2006/7 (n=58) | Sentinel study 2015-2018 (n=535) |
|---|---|---|---|---|
| Aminoglycoside | 1 (6%) | 5 (14%) | 7 (12%) | 64 (12%) |
| Fluoroquinolone | 1 (6%) | 14 (38%) | 16 (28%) | 196 (37%) |
| Macrolide | 0 (0%) | 1 (3%) | 6 (10%) | 28 (5%) |
| Tetracycline | 1 (6%) | 13 (35%) | 25 (43%) | 204 (38%) |

Table 61. Antimicrobial resistance among *Campylobacter coli* from food animal isolates sequenced as part of FS101013.

| | Chicken (abattoir) (n=221) | Chicken (retail) (n=108) | Ox/calf liver (n=20) | Lamb liver (n=7) | Duck (n=11) | Pigs (n=83) |
|---|---|---|---|---|---|---|
| Aminoglycoside | 26 (12%) | 14 (13%) | 2 (10%) | 1 (14%) | 2 (18%) | 49 (59%) |
| Fluoroquinolone | 112 (51%) | 52 (48%) | 4 (20%) | 1 (14%) | 6 (55%) | 14 (17%) |
| Macrolide | 3 (1%) | 2 (2%) | 0 (0%) | 0 (0%) | 0 (0%) | 28 (34%) |
| Tetracycline | 141 (64%) | 65 (60%) | 2 (10%) | 1 (14%) | 9 (82%) | 71 (86%) |

## Results from application of BIGSdb to predict antimicrobial resistance

A bar graph showing the proportion of isolates (%) resistant to the defined antimicrobials across 6 sampling years.

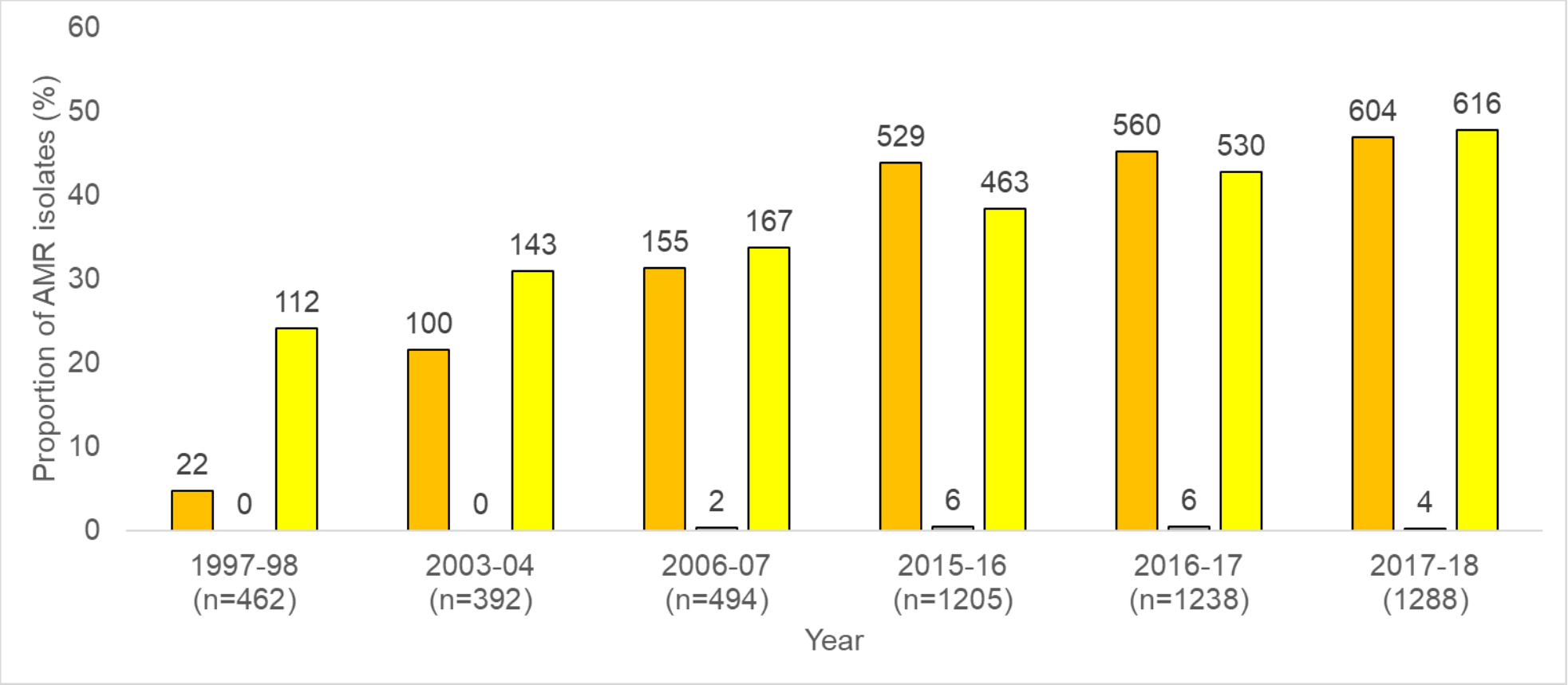



Figure 49. Percentage of fluoroquinolone (orange), macrolide (grey) and tetracycline (yellow) resistant *C. jejuni* clinical isolates, as determined from WGS analysis by BIGSdb. Numbers above each column indicate the number of resistant isolates.

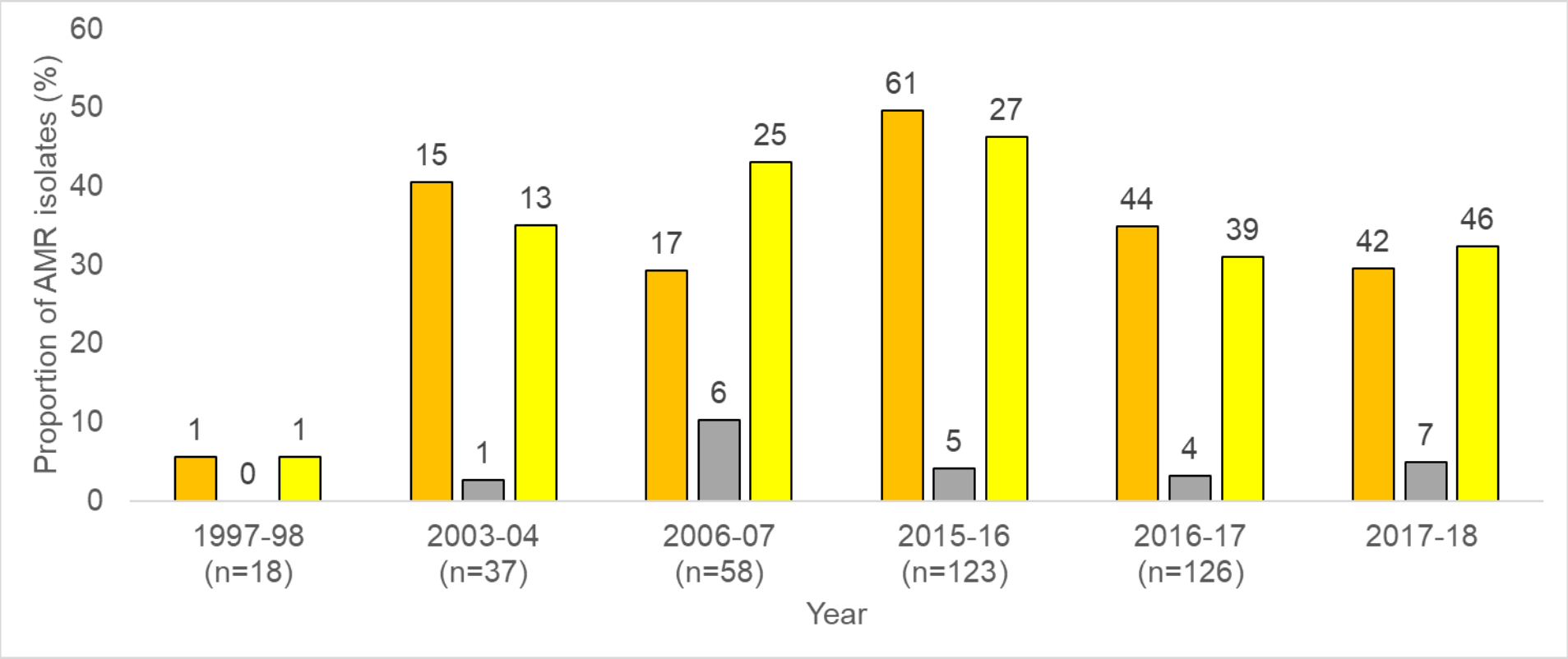



Figure 50. Percentage of fluoroquinolone (orange), macrolide (grey) and tetracycline (yellow) resistant *C. coli* clinical isolates, as determined from WGS analysis by BIGSdb. Numbers above each column indicate the number of resistant isolates.

Macrolide resistance levels among clinical *C. jejuni* are lower than for *C. coli*, with maximum values seen between 2015 and 2017 of 0.5%. Maximum FQ resistance values have shown a steady increase in *C. jejuni* from 4.8% in 1997-98 to 46.9% in 2017-18. Tetracycline resistance has also increased from 24.2% in 1997-98 to 47.8% in 2017-18. Maximum resistance levels for FQ and tetracycline are similar for both species. Among *C. coli* AMR for all three antibiotics was low in the epidemiological year 1997-98 but increased to 40.5%, 2.7% and 35.1% for FQ, macrolide and tetracycline resistance respectively by 2003-04. Whilst there was a decline in the proportion of FQ resistant isolates to 29.3% in 2006-07, macrolide and tetracycline resistance continued to increase to 10.3% and 43.1%. FQ and tetracycline resistance was highest in 2015-16 (49.6% and 46.3%), with a slight decline to 4.1% for macrolides. Subsequently there have been slight fluctuations in levels of resistance for all three antibiotic classes, which all remain lower than in 2015-16.

Fluoroquinolone resistance was highest *C. jejuni* isolates from the retail chicken (47.6%) and chicken abattoir (42.2%) surveys. Beef offal, duck and lamb offal isolates all had fluoroquinolone resistance levels of below 15% (13.1%, 9.4% and 7.4% respectively). No macrolide resistant isolates were detected among abattoir or food *C. jejuni* isolates from any source. Tetracycline resistance was highest in *C. jejuni* isolated from the pig abattoir survey (66.7%), duck (59.4%) and retail (57%) and abattoir chicken (50.6%) survey samples.

Fluoroquinolone resistance was highest among *C. coli* isolates from beef offal at 57.1% and duck isolates at 54.5%, followed closely by retail chicken at 48.8%, however the number of *C. coli* isolates available from beef and duck were small (20 and 11 respectively). Macrolide resistance was highest among pig isolates (21.7%) with low level resistance in retail chicken meat (1.9%). No macrolide resistance was detected among chicken abattoir, duck and lamb or beef offal. Tetracycline resistance was highest among pig isolates (79.7%), followed by duck (72.7%) retail chicken (61.5%) and chicken abattoir (61.3%). Lower values were observed for beef (28.6%) and lamb offal (14.3%).

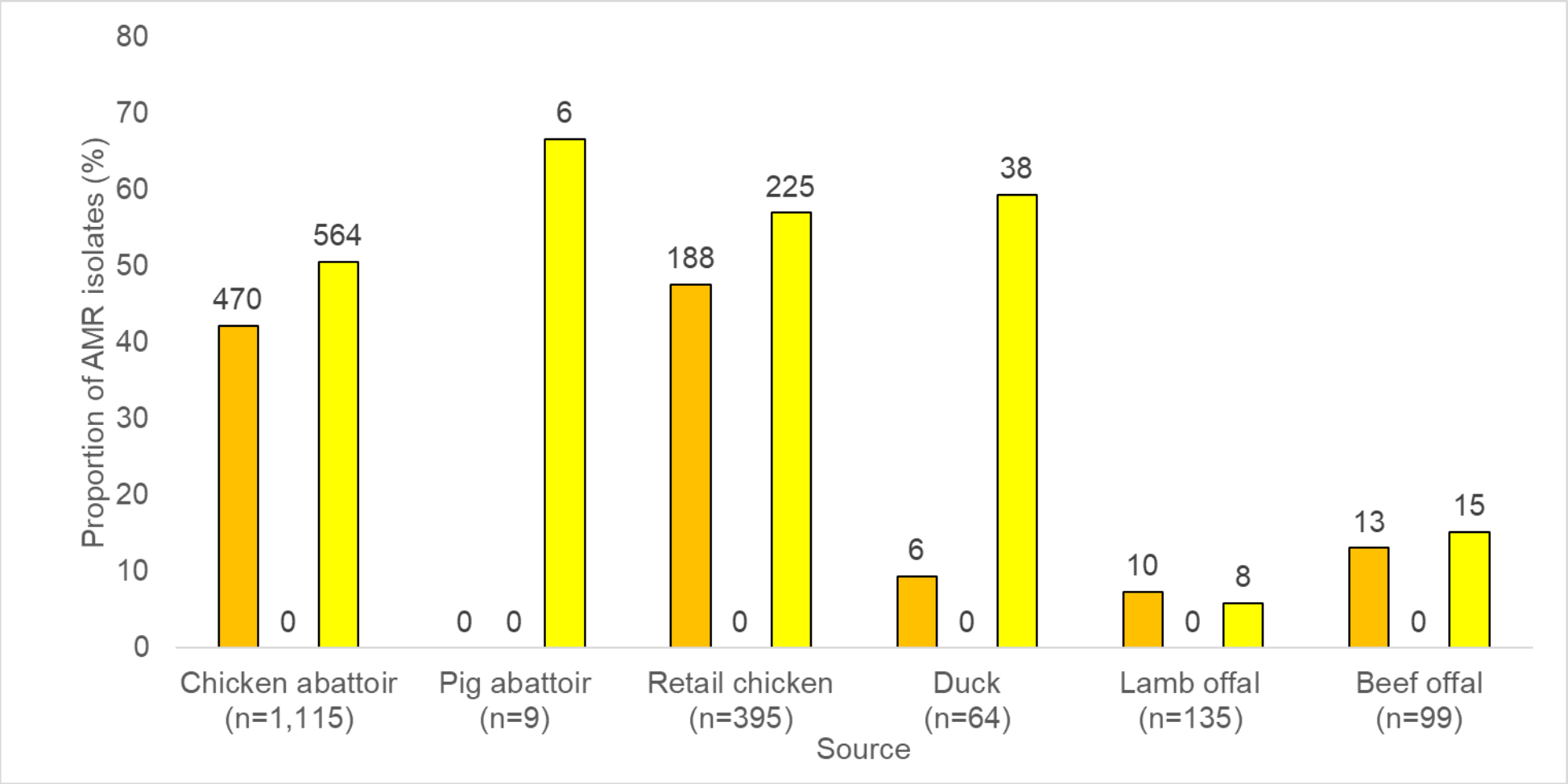



Figure 51. Percentage of fluoroquinolone (orange), macrolide (grey) and tetracycline (yellow) resistant *C. jejuni* among chicken abattoir, pig abattoir and retail meat isolates, as determined from WGS analysis by the BIGSdb. Numbers above each column indicate the number of resistant isolates.

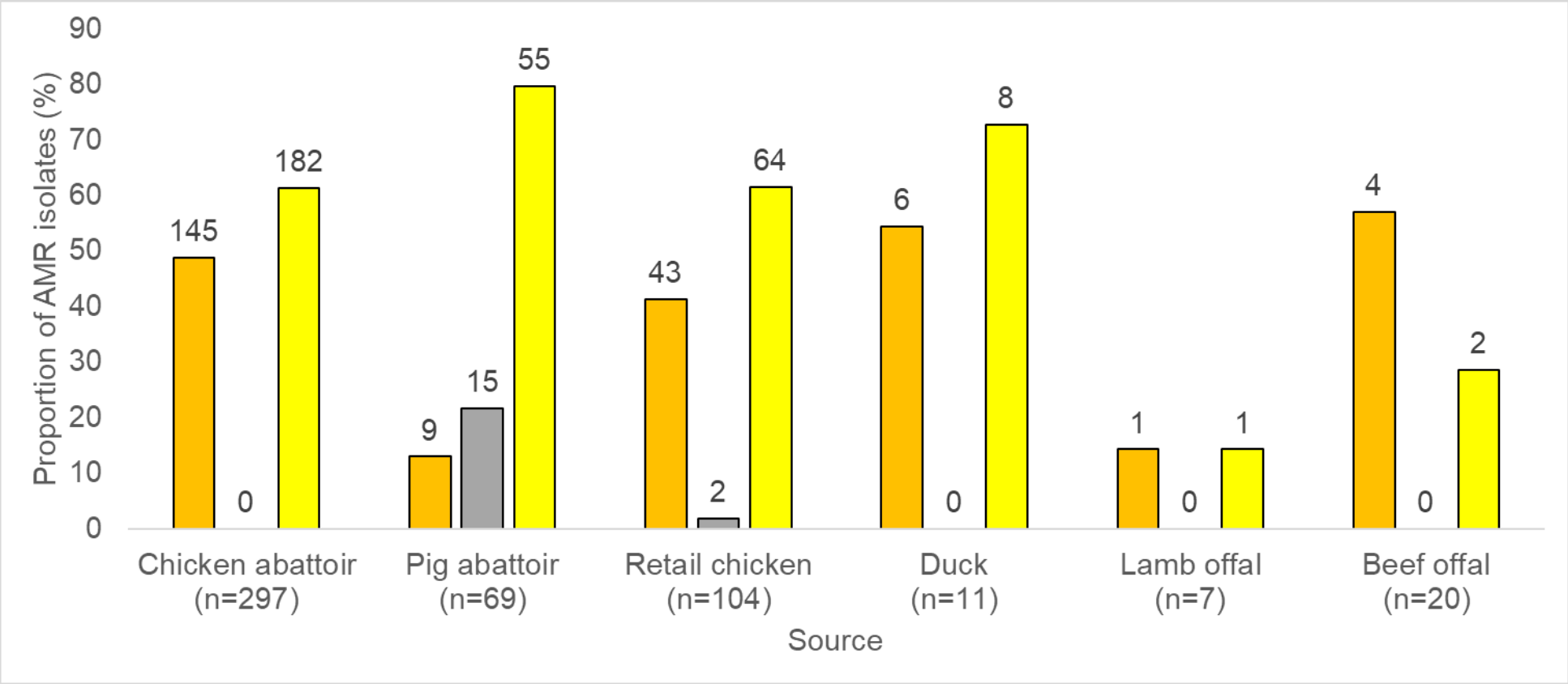



Figure 52. Percentage of fluoroquinolone (orange), macrolide (grey) and tetracycline (yellow) resistant *C. coli* among chicken abattoir, pig abattoir and retail meat isolates, as determined from WGS analysis by the BIGSdb Numbers above each column indicate the number of resistant isolates.

## 13.10 Appendix 10 Glossary and abbreviations

Apparently sporadic – cases of a disease with no evidence for their being lined epidemiologically to other cases as part of an outbreak.

Assembly statistics – summary statistics from genome assembly, such as the number of fragments and overall estimated genome size.

Attribution – estimating the proportions of infection from different sources, "population genetic attribution" when this is done using population genetic analysis algorithms on microbial genetic data.

Clonal Complex (CC) – the grouping of sequence types that share similarity with a central allelic profile.

Contigs (contiguous) – a set of overlapping DNA segments that together represent a consensus region of DNA.

iSource The software implementation of the asymmetric island model for attribution by Daniel Wilson available here, visited 18 August 2020

Modified Charcoal-Cefoperazone-Deoxycholate Agar (mCCDA) – selective plating medium used for the detection and enumeration of *Campylobacter spp.*

Multi-locus Sequence Typing (MLST) – a technique used in molecular biology for the typing of bacteria based on genetic sequencing of multiple loci, typically seven housekeeping genes unless otherwise specified. Other specified variations of this include core genome Multi-locus Sequence Typing (cgMLST) – a similar approach using all loci which are present in all or almost all strains of that species and whole genome MLST (wgMLST) which uses all genes known to be present in the bacterial species.

Paralogous loci – genetic loci at difference chromosomal locations that have structural similarities suggesting they are from a common ancestral gene and diverged by mutation and selection or drift.

PubMLST – a publicly accessible database for molecular typing using MLST.

Sequence Type (ST) – a type defined by the profile of alleles at each of the loci within the MLST scheme.

Single-linkage clustering (SLC) – a method of hierarchical clustering based on grouping clusters in a bottom-up fashion, combining two isolates and clusters that contain a close pair of elements.

Single-Nucleotide Polymorphisms (SNPs) – a substitution of a single nucleotide at a specific position in the genome.

BIGSdb - Bacterial Isolate Genome Sequence Database

cfu – Colony Forming Unit

ENA – European Nucleotide Archive

FSA – Food Standards Agency

mCCDA - Modified Charcoal-Cefoperazone-Deoxycholate Agar

WGS – Whole Genome Sequencing.