Investigating locally relevant risk factors for *Campylobacter* infection in Australia: protocol for a case–control study and genomic analysis

Liana Varrone, Russell J Stafford, Kim Lilly, Linda Selvey, Kathryn Glass, Laura Ford, Dieter Bulach, Martyn D Kirk, On behalf of The CampySource Project Team

ABSTRACT

**Introduction** The CampySource project aims to identify risk factors for human *Campylobacter* infection in Australia. We will investigate locally relevant risk factors and those significant in international studies in a case–control study. Case isolates and contemporaneous isolates from food and animal sources will be sequenced to conduct source attribution modelling, and findings will be combined with the case–control study in a source-assigned analysis.

**Methods and analysis** The case–control study will include 1200 participants (600 cases and 600 controls) across three regions in Australia. Cases will be recruited from campylobacteriosis notifications to health departments. Only those with a pure and viable *Campylobacter* isolate will be eligible for selection to allow for whole genome sequencing of isolates. Controls will be recruited from notified cases of influenza, frequency matched by sex, age group and geographical area of residence. All participants will be interviewed by trained telephone interviewers using a piloted questionnaire. We will collect *Campylobacter* isolates from retail meats and companion animals (specifically dogs), and all food, animal and human isolates will undergo whole genome sequencing. We will use sequence data to estimate the proportion of human infections that can be attributed to animal and food reservoirs (source attribution modelling), and to identify spatial clusters and temporal trends. Source-assigned analysis of the case–control study data will also be conducted where cases are grouped according to attributed sources.

**Ethics and dissemination** Human and animal ethics have been approved. Genomic data will be published in online archives accompanied by basic metadata. We anticipate several publications to come from this study.

INTRODUCTION

*Campylobacter* infection is the most commonly notified cause of foodborne gastroenteritis in Australia, as well as a leading cause of bacterial gastroenteritis worldwide. At the introduction of Australia’s National Notifiable Diseases Surveillance System in 1991 the incidence rate of notified campylobacteriosis cases was 79.1/100 000 population, and despite notification rates plateauing in recent years, incidence had risen to 139.7/100 000 population in Australia in 2015, with an estimated 10 cases for every notified case within the community. Similarly, the incidence rate of campylobacteriosis in New Zealand in 2014 was 150.3/100 000 population, with an estimated 10–30 cases in the community for every notified case. *Campylobacter* notification rates in Australia and New Zealand are still among the highest in the world across high-income countries. Most countries in the European Union consistently report annual campylobacteriosis notification rates below 100/100 000 population.

Two species of *Campylobacter*—*Campylobacter jejuni* and *C. coli*—contribute to approximately 95% of human campylobacteriosis. These *Campylobacter* spp are commonly detected in sewage and surface water, reside in the gastrointestinal tract of birds and animals and are frequently found in raw meat, particularly poultry, and raw milk. Campylobacteriosis is mostly foodborne, with an
estimated 77% of cases transmitted via food consumption in Australia. Direct and indirect zoonotic transmission can occur via animal contact (direct) or faecally contaminated water or environments (indirect). Person-to-person transmission is considered rare. The majority of cases are thought to be sporadic, with outbreaks less commonly detected. Most outbreaks are linked to the consumption of poultry, raw milk or contaminated water. Targeted control of foodborne bacterial pathogens generally depends on identification of sources and routes of transmission. Since *Campylobacter* are ubiquitous in the environment and most cases are sporadic, identifying sources is difficult. Source attribution methods require isolation of strains from reservoirs to compare *Campylobacter* strain diversity in foods and animals with that in human infections. Beef, sheep and pig meat have a lower prevalence of *Campylobacter* contamination than chicken meat (<5%–14%), but a higher prevalence is found in animal offal such as liver, thus making offal a valuable source of host-associated strains of *Campylobacter* in low-prevalence meats.

**STUDY RATIONALE**

In the USA, evidence from case–control studies has led to policy change, including changes to chicken slaughtering techniques. The incidence of human *Campylobacter* infection has declined in the USA since this policy was introduced in 1997. More recently, evidence from source attribution analyses in New Zealand has led to the development of poultry production policies and practices aimed at reducing the risk of *Campylobacter* transmission via poultry food products. New Zealand has seen a 74% reduction in the number of campylobacteriosis cases attributed to poultry in the region, as well as a 54% reduction in cases overall.

Source attribution modelling enables us to determine which foods and animals are the most likely sources of infection with each *Campylobacter* strain type, and the proportion of cases attributed to each source. This can be done with simple proportional similarity index (PSI) calculations, or by using more complex models. Source attribution also allows for human campylobacteriosis cases to be grouped by potential source, increasing the specificity of risk factor analyses. These source-assigned analyses combine the epidemiological information gained through the traditional case–control study with source attribution modelling to provide greater explanatory power to investigate locally relevant risk factors.

**OBJECTIVES**

This study aims to:

1. Identify dietary, environmental and behavioural risk factors for *Campylobacter* infection in Australia.
2. Strengthen the epidemiological evidence for previously identified risk factors in Australia.
3. Identify strain-specific risk factors for infection using whole genome sequencing (WGS) data from case isolates.

**HYPOTHESES**

We will test several hypotheses regarding specific risk factors for *Campylobacter* infection in Australia. The hypotheses are based on exposures that have previously been identified as risk factors for *Campylobacter* infection in Australia as well as internationally.

We hypothesise that:

1. Persons who consume undercooked meats, particularly chicken, are at increased risk of infection.
2. Persons who consume offal are at increased risk of infection.
3. Persons who own companion animals (especially puppies) are at increased risk of infection.
4. Poor food hygiene and handling practices in the home increase the risk of infection.
5. Most human infections will be attributed to consumption of chicken meat.
6. There will be a high level of genetic diversity among *Campylobacter* strains.

**STUDY DESIGN**

We will conduct a case–control study including genomic testing over a 2-year period in three sentinel sites: the state of Queensland (QLD), the Australian Capital Territory (ACT) and the Hunter New England (HNE) region of New South Wales (figure 1). Sporadic cases of culture-positive *Campylobacter* infection will be identified either through state notifiable disease registers, from local pathology service databases or local notification databases. An isolate from each case will be paired with epidemiological data from the case interview. One control will be recruited for each case who participates in the study, with trained interviewers conducting telephone interviews with both cases and controls.
controls. Participants will be interviewed using a questionnaire that has been specifically designed to collect information on known potential risk factors. This questionnaire will include a selection of questions being validated in a separate study (LV: Validation of questions designed for investigation of gastroenteritis). For cases, the questions will cover the 7 days prior to the onset of illness, while controls will be questioned on the 7 days prior to interview. Meanwhile, Campylobacter isolates will also be collected from food and animal samples. All human and non-human isolates will undergo WGS for comparison in source attribution modelling. Data for this study will be collected from 1 March 2017 to 1 March 2019.

Patient and public involvement

To develop the study, we engaged state and territory health departments, food safety agencies and industry to establish research questions and methods. The process involved a dedicated workshop, followed by teleconferences and an iterative process of drafting study documentation. We also established a reference panel, which includes representatives from senior levels of government and industry bodies. No patients or other members of the public were involved in the development of this study.

### STUDY POPULATION

The three sentinel sites cover a population of approximately 6.1 million people. Based on notification and diagnostic pathology data, we expect approximately 8650 *Campylobacter* cases to be notified across these sites during the study period.

### DEFINITION AND SELECTION OF CASES

#### Case definition

We define a case as a person from any of the three participating sites with a history of acute diarrhoea and a culture-positive stool result for *Campylobacter*.

### SAMPLE SIZE

We used risk factor prevalence data from a previous national *Campylobacter* case–control study in 2001/2002 to estimate sample size for this study. For example, the prevalence of chicken consumption among controls in 2001/2002 was 80%. A sample size of approximately 1040 subjects (520 cases; 520 controls) would enable the study to detect an association between chicken consumption and illness with an OR of 1.6, at 80% power and α=0.05, as reported in the previous study. Sample size estimates for other potential risk factors are listed in table 1.

From these calculations, we estimate that a study of 1200 subjects (600 cases; 600 controls) will adequately detect significant associations of these magnitudes for potential risk factors of interest. QLD and HNE sites will each enrol at least 250 cases into the study, while ACT will enrol at least 100 cases. Based on the previous Australian case–control study, we expect approximately 80% of selected notified cases to be eligible and participate in the study (table 2).

In QLD, we will obtain cases from one private pathology provider reporting approximately 40% of the state’s *Campylobacter* notifications. We estimate that this provider will notify 2800 cases during the study period with an estimated 45% of these being culture positive (1260 notified cases). In ACT, approximately 600 *Campylobacter* notifications are expected during the study period; 130 are expected from the participating pathology laboratory in...
Table 3  Eligibility criteria for cases and controls

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Had diarrhoea (≥3 loose bowel movements in 24 hours)</td>
<td>Include</td>
<td>Exclude</td>
</tr>
<tr>
<td>Known date of illness onset</td>
<td>Include</td>
<td>NA</td>
</tr>
<tr>
<td>Household members positive for <em>Campylobacter</em> in 4 weeks prior to onset of illness</td>
<td>Exclude</td>
<td>Exclude (4 weeks prior to interview date)</td>
</tr>
<tr>
<td>Household members experiencing diarrhoea in 4 weeks prior to onset of illness</td>
<td>Exclude</td>
<td>Exclude (4 weeks prior to interview date)</td>
</tr>
<tr>
<td>Travelled outside of Australia in 2 weeks prior to onset of illness</td>
<td>Exclude</td>
<td>Exclude (2 weeks prior to interview date)</td>
</tr>
<tr>
<td>Travelled interstate for the entire 2 weeks prior to onset of illness</td>
<td>Exclude</td>
<td>Exclude (2 weeks prior to interview date)</td>
</tr>
<tr>
<td>Cannot speak English</td>
<td>Exclude</td>
<td>Exclude</td>
</tr>
<tr>
<td>Not able to answer questions for some other reason (eg, intellectually disabled)</td>
<td>Exclude</td>
<td>Exclude</td>
</tr>
<tr>
<td>Not contactable after six telephone attempts</td>
<td>Exclude</td>
<td>Exclude</td>
</tr>
<tr>
<td>Live outside the catchment areas</td>
<td>Exclude</td>
<td>Exclude</td>
</tr>
<tr>
<td>Do not have a telephone number available for their primary residence, or a mobile phone</td>
<td>Exclude</td>
<td>Exclude</td>
</tr>
<tr>
<td>An enteric pathogen other than <em>Campylobacter</em> was isolated/detected in their stool (excluding Blastocystis hominis and Dientamoeba fragilis)</td>
<td>Exclude</td>
<td>NA</td>
</tr>
</tbody>
</table>

NA, not applicable.

HNE, approximately 1050 *Campylobacter* notifications are expected to be from the participating pathology laboratory.

Enrolment of cases

We will enrol all cases who meet the eligibility criteria (table 3). Each site will check for new notifications of culture-positive *Campylobacter* infection daily, with only culture-positive *Campylobacter* cases eligible for this study. If a case refuses to participate in the study, we will select a subsequent case for inclusion. Enrolment of cases will require consent from the patient, or in the event of a child aged less than 18 years, consent from either one of the parents or the child’s guardian. We will interview cases as soon as possible by telephone, preferably within 2 weeks of notification from the laboratory. It will be at the parent’s or guardian’s discretion as to whether a child aged between 15 and 17 years is interviewed directly. The parent or guardian will be interviewed for cases aged less than 15 years.

**DEFINITION AND SELECTION OF CONTROLS**

We will recruit controls from notified cases of influenza, frequency matched by sex, age group and geographical area of residence by statistical area level 4 (SA4). These controls will be selected with a delay of at least 6 months from their influenza infection to ensure that controls have returned to eating their customary diet.

Each participating site (QLD, ACT or HNE) will establish a database of controls (previous influenza cases). All cases of influenza notified to the health department in each site between 1 January and 31 December 2017 will be entered into this control database. The age bands are 0–4 years, 5–14 years, 15–34 years, 35–54 years, 55–74 years and ≥75 years. An appropriate control will be randomly selected from the database within 30 days of interview of the notified case.

**Case and control recruitment**

Interviewers trained in computer-assisted telephone interviewing will conduct telephone interviews. A maximum of six attempts will be made to contact any one case or control, with no more than three attempts in any one day. Three calls will be attempted between 09:00 and 15:59, and three attempts between 16:00 and 20:00. A text message will be sent to the potential participant after three failed call attempts, indicating that Public Health is trying to contact them. This protocol will be continued until the person is enrolled or excluded.

**QUESTIONNAIRES**

We will use specific case and control questionnaires for all participants (see online supplementary appendix 1). Cases will be asked additional questions about the clinical course of their illness and treatment. Interviewers will ask identical questions regarding exposures such as foods consumed, dining locations, water sources, domestic food handling techniques and exposure to animals of cases and controls. Questions on foods consumed, dining locations, water consumed and animal and pet exposures will be asked based on a 7-day history. Questions on international travel will be asked based on a 2-week history. Antibiotic and antacid consumption, immunosuppressive treatment and household history of diarrhoea will be based on a 4-week history. Questions on food handling and general kitchen practices will be based on usual

practices rather than recent history. Demographic information will be collected from cases and controls. Contact information required to conduct interviews will be stored in a password-protected Excel document with only those needing to contact individuals given access. Piloted questionnaires were modified to remove repetitions, improve clarity and to ensure that interviews could be conducted within 20 min.

**DATA HANDLING AND RISK FACTOR ANALYSIS**

We will undertake descriptive reporting of campylobacteriosis incidence by person, place and time. We will also describe the severity of symptoms, treatment and burden of illness.

Risk factor analysis will involve the examination of 2×2 contingency tables with χ² or exact tests to determine the presence of univariable associations between variables and disease. To measure the strength of an association, we will estimate ORs and calculate 95% CIs in a univariable analysis, followed by multivariable logistic regression modelling to adjust for potential confounders. Risk factors selected for inclusion in the regression model will include age, season and geographic area, variables with a significant univariable association with disease, and variables with a p value ≤0.25 that are biologically plausible and of interest to the research team.

**LABORATORY ANALYSES**

**Human samples**

As outlined in table 2, it is expected that 250 human isolates from HNE, 250 from QLD and 100 from ACT will be sequenced, with an additional 100 isolates being sourced from Victoria. The initial isolation and confirmation of *Campylobacter* infection will be performed locally in each state/territory. Only samples with a pure and viable culture will undergo WGS.

We will also collect an additional 20 – 30 human isolates from four Australian jurisdictions not participating in this case – control study to undergo WGS. This will be done over a 2-month period that overlaps with the case – control study sample collection, and is planned to help inform the generalisability of the case – control study.

**Animal and food samples**

We will collect samples from chicken meat (covering the two production methods of continually housed and free range/housed), beef, lamb, pork and pet dogs. Given low prevalence of *Campylobacter* in meats other than chicken, samples will be collected from offal (preferably liver) from bovine, ovine and porcine sources to ensure sufficient positive samples are obtained for the study. Given the rising importance of chicken liver pate as a source of outbreaks in Australia, 27 chicken offal will also be sampled. Sample sizes by source are based on data from two states to ensure 50 positive samples per food source, and 30 samples in companion animals (table 4). We will also contact veterinary clinics and teaching hospitals to ensure sufficient *Campylobacter*-positive samples from dogs. Water samples have been omitted from the genomic aspect of this study due to logistical constraints in sampling untreated water sources across the large geographical area involved in this study, and the complexity of designing an appropriate sampling frame. As there is a lack of evidence implicating municipal drinking water as sources of *Campylobacter* infection in Australia, 1 26 we excluded water sampling from this study.

The initial isolation and confirmation of *Campylobacter* will be performed locally at laboratories in each state/territory, with isolates forwarded to the Microbiological Diagnostic Unit Public Health Laboratory for WGS, except QLD isolates which will be sequenced at Queensland Health. To detect seasonal and temporal variation in *Campylobacter* genetic types, 1041 food and animal samples (estimated to produce 330 *Campylobacter* isolates) will be collected over a period of 1 year in QLD, and 2 years in New South Wales. To assess latitudinal variation in chicken meat samples across eastern Australia, 105 chicken samples (70 chicken meat and 35 chicken offal) will be collected over a 6-month period in Victoria. Food samples will be collected monthly from retail premises, using protocols from surveys undertaken in 2014 by partner organisations, with a pilot of 30 isolates in QLD.

**SEQUENCING AND SEQUENCE DATA PROCESSING**

*Campylobacter* isolates selected for sequencing will be repurified on solid medium and a single colony selected for preparation of genomic DNA. A sequencing library will be prepared from the genomic DNA for sequencing on the Illumina sequencing platform (MiSeq or NextSeq). A sample of the selected colony will be regrown and cryopreserved (resuspended in liquid medium supplemented with 10% glycerol and stored at −80°C). In some cases,

<table>
<thead>
<tr>
<th>Foods</th>
<th>Chicken</th>
<th>Free range</th>
<th>Offal</th>
<th>Beef</th>
<th>Lamb</th>
<th>Pork</th>
<th>Dogs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed prevalence</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.14</td>
<td>0.6</td>
<td>0.22</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Samples required</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>286</td>
<td>100</td>
<td>272</td>
<td>150</td>
<td>1041</td>
</tr>
<tr>
<td>Positive isolates</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>40</td>
<td>60</td>
<td>60</td>
<td>30</td>
<td>330</td>
</tr>
</tbody>
</table>

Table 4 Sampling to ensure 50 isolates per food source and 30 isolates from companion animals

Campylobacter enrichment cultures will be cryopreserved to enable future investigation of the genetic diversity of Campylobacter present. The short-read, paired-end data set produced by the Illumina Instrument from the genomic DNA of each isolate will be processed to produce a draft genome sequence for the isolate using a de novo assembler such as MEGAHIT. The draft genome sequence will be annotated using Prokka. We will use the draft genome sequence to perform the initial subspecies classification by deriving a multilocus sequence type (MLST) using the ‘Campylobacter jejuni/coli’ typing scheme (pubmlst.org). Again, using the draft genome sequence, further typing, for example, virulence factors (http://www.mgc.ac.cn/VFs/) or antimicrobial resistance genotype (https://cge.cbs.dtu.dk/services/ResFinder/), will be performed using Abricate (https://github.com/tseemann/abricate). We will perform comparative genomics to examine the genetic relationships between selected subgroups of isolates in more detail using Nullarbor (https://github.com/tseemann/nullarbor).

SOURCE ATTRIBUTION MODELLING
We will analyse the epidemiological data within designated MLST groups or other typing groups derived from the genomic sequence data. Source attribution modelling and source-assigned analyses will be conducted.

Source attribution models combine typing data from isolates from food, animals and humans to estimate the proportion of human infections that can be attributed to animal and food reservoirs. Once inferred MLSTs have been ascertained, the PST will be used to assess similarities by source. We will then undertake source attribution analyses by adapting the asymmetric island model which has previously been applied to MLST data using Markov chain Monte Carlo methods implemented using the free software WinBUGS. These methods will first be applied to MLST data extracted from whole genome sequences (the aforementioned ‘inferred MLSTs’), and then compared with structured phylogenetic modelling approaches that provide scope to infer interhost transmission.

We will then group cases according to putative source based on these source attribution methods. For example, all isolates attributed to chicken will be grouped together, regardless of differing strains. These cases attributed to chicken will then be compared with all controls in a risk factor analysis to produce a source-assigned analysis.

SPATIAL CLUSTERS AND TEMPORAL TRENDS
We will use newly designated WGS-based MLSTs to assess heterogeneity in isolates from food sources and companion animals in QLD and New South Wales, and in isolates from chicken meat and humans across QLD, New South Wales, Victoria and ACT. A 2-year sampling framework in New South Wales, 1 year of sampling in QLD and previous survey work in these states will allow us to assess the extent of seasonal and temporal trends. Postcode-level data associated with human illnesses will be used to detect space-time clusters using a scan statistic implemented in the free software SaTScan, at the SA1. We will use a retrospective space-time permutation model to detect high-risk clusters by comparing the observed number of illnesses with the expected number in that geographic zone and time period.

Author affiliations
1Faculty of Medicine, University of Queensland, Herston, Queensland, Australia
2Communicable Diseases Branch, Queensland Health, Brisbane, Queensland, Australia
3Hunter New England Population Health, Newcastle, New South Wales, Australia
4National Centre for Epidemiology and Population Health, The Australian National University, Canberra, Australian Capital Territory, Australia
5College of Health and Medicine, Melbourne Bioinformatics, Melbourne, Victoria, Australia
6Microbiological Diagnostic Unit Public Health Laboratory, The Peter Doherty Institute, Melbourne, Victoria, Australia

Collaborators Study linkages and collaborations: The CampySource Project Team comprises three working groups and a reference panel. The working groups focus on: food and animal sampling, epidemiology and modelling, and genomics. The reference panel includes expert representatives from government and industry. The study is supported by the following partner organisations: the Australian National University, Massey University, University of Melbourne, Queensland Health, Queensland Health Forensic and Scientific Services, New South Wales Health, Hunter New England Health, Victorian Department of Health and Human Services, Food Standards Australia New Zealand, Commonwealth Department of Health and AgriFutures Australia—Chicken Meat Program. CampySource is also supported by collaboration with the following organisations: ACT Health, Sullivan Nicolaides Pathology, University of Queensland, Primary Industries and Regions South Australia, Department of Health and Human Services Tasmania, Meat and Livestock Australia, and New Zealand Ministry for Primary Industries. The CampySource Project Team consists of: Nigel P French, Massey University, New Zealand; Mary Valcakis, The University of Melbourne; Dieter Baluch, The University of Melbourne; Emily Fearnley, The Australian National University; Russell Stafford, Queensland Health; John Bates, Queensland Health; Trudy Graham, Queensland Health; Keira Glasgow, Health Protection NSW; Kirsty Hope, Health Protection NSW; Arie H Havelaar, The University of Florida, USA; Joy Gregory, Department of Health and Human Services, Victoria; James Flint, Hunter New England Health; Simon Firestone, The University of Melbourne; James Conlan, Food Standards Australia New Zealand; James J Smith, Queensland Health; Sally Symes, Department of Health and Human Services, Victoria; Barbara Butow, Food Standards Australia New Zealand; Lisa Varrone, The University of Queensland; Linda Selvey, The University of Queensland; Deborah Dench, ACT Health; Radomir Krsteski, ACT Health; Natasha Waters, ACT Health; Kim Lilly, Hunter New England Health; Julie Collins, Hunter New England Health; Tony Merritt, Hunter New England Health; Joanne Barfield, Hunter New England Health; Ben Howden, The University of Melbourne; Kylie Hewson, AgriFutures Australia—Chicken Meat Program; Laura Ford, The Australian National University; Liz Walker, The Australian National University; Cameron Moffatt, The Australian National University; Martyn Kirk, The Australian National University; Kathryn Glass, The Australian National University.

Contributors MDK conceived the original idea for this study. All authors contributed to the study design and analysis plan. LV and RJS wrote the first draft with contributions from all authors. LF was heavily involved in determining timing and logistics in and between all sites. KL assisted in questionnaire design and flow. DB developed the bioinformatics analysis protocol. LV, RJS, LS, MDK and KG were involved in multiple revisions. The final version of the manuscript was approved by all authors.

Funding This CampySource project was supported by an NHMRC partnership grant 1116294 and contributions from Queensland Health, Food Standards Australia New Zealand, AgriFutures Australia—Chicken Meat Program, Commonwealth Department of Health, and Queensland Health Department of Primary Industries. MDK is supported by a National Health and Medical Research Council Fellowship (GNT1145997). While undertaking studies, LV is supported through an Australian Government Research Training Program (RTP) Scholarship.