Freshwater Microbiological Research Programme 2020-2025

July 2025 (Updated June 2026 with minor corrections)

Report Authors:

Margaret Leonard (PHF Science)
Carla Eaton (PHF Science)
Brent Gilpin (PHF Science)
Beverley Horn (PHF Science)
David Wood (Earth Sciences NZ)

Project Team:

Angela Cornelius (PHF Science)
Susan Lin (PHF Science)
Sarah Coxon (PHF Science)
Paula Scholes (PHF Science)
Kirsten Thom (PHF Science)
Kathryn Russell (PHF Science)
Anthony Pita (Massey University)
Beth Robson (PHF Science)
Helen Morris (PHF Science)
Bridget Armstrong (PHF Science)
Iain Haysom (PHF Science)
Tori Head (PHF Science)
Joanne Hewitt (PHF Science)
David Hayman (Massey University)

PREPARED FOR: **Ministry for the Environment**

CLIENT REPORT No: **FW25002**

PHF Science Report Reviewers:

Meg Devane (PHF Science) Peter Cressey (PHF Science)

Draft Report Reviewers:

Richard Muirhead (Bioeconomy NZ)

Two Anonymous reviewers based in the USA

ACKNOWLEDGEMENTS

We would like to acknowledge and thank the staff from the following Regional and Unitary councils for their time and efforts in undertaking sample collection for this study, braving some unpleasant weather and frustrations with courier services and the Council managers who have supported this project in principle and made the resources available.

- Auckland Council
- Bay of Plenty Regional Council
- Environment Canterbury
- Environment Southland
- Greater Wellington Regional Council
- Gisborne District Council
- Hawkes Bay Regional Council
- Horizons Regional Council
- Marlborough District Council
- Nelson City Council
- Northland Regional Council
- Taranaki Regional Council
- Tasman District Council
- Waikato Regional Council
- West Coast Regional Council.

We would also like to thank iwi and hapū who generously gave their time and knowledge of their awa.

We acknowledge the valuable contributions from the Governance and Stakeholder Groups

Governance Group: Tim Davie (Environment Canterbury), Tim Hopley (Ministry of Health), Carl Howarth (Ministry for the Environment), Gillian Lewis (University of Auckland), Richard Muirhead (AgResearch), Jeff Soller (Soller Environmental), Winona Bea Molano (Ministry for the Environment).

Stakeholder Group: Winona Bea Molano (Ministry for the Environment), Tim Hopley (Te Whatu Ora), Carl Howarth (Ministry for the Environment), Abi Loughnan (LAWA), David Wood (NIWA), Mathirimangalam Srinivasan (Ministry for Primary Industries), David Sinclair (Auckland District Health Board), Coral Grant (Auckland Council), Craig Depree (Dairy NZ), Mike Scarsbrook (Waikato Regional Council), Margaret Foster, David Acland.

We would like to thank Juliet Milne, NIWA, who liaised with the Councils to select suitable sites for the Pilot study and prepared the contact list and Elena Moltchanova, University of Canterbury for useful discussions on quantile regression modelling

We would also like to acknowledge our internal and external reviewers and colleagues for their comments and insight. Our thanks to the laboratory staff that did a great job managing the logistics as well as the analyses.

CONTENTS

1 INTRODUCTION 8
1.1 BACKGROUND 8
1.2 RATIONALE FOR A NEW SURVEY AND QMRA 9
1.3 STUDY OUTLINE 10
1.4 AIMS OF THE STUDY 11
1.5 REPORT OUTLINE 11
2 FRESHWATER SURVEY INTRODUCTION 14
2.1 INDICATORS OF FAECAL POLLUTION 14
2.1.1 Total coliforms, *E. coli* and enterococci 14
2.1.2 Microbial source tracking markers 15
2.2 PATHOGENS 16
3 SITES AND SAMPLING DESIGN 19
3.1 SITE SELECTION 19
3.2 OBSERVED LAND USE CATEGORY 21
3.3 SAMPLES COLLECTED 21
4 FAECAL INDICATOR BACTERIA 22
4.1 CULTURE-BASED ANALYSIS OF TOTAL COLIFORMS, *E. COLI* AND ENTEROCOCCI 22
4.2 QPCR-BASED ANALYSIS OF *E. COLI* AND ENTEROCOCCI 26
5 SITE CLASSIFICATION BY MICROBIAL SOURCE TRACKING 28
5.1 MICROBIAL SOURCE TRACKING ANALYSES 28
5.2 CATEGORISATION BY MICROBIAL SOURCE TRACKING MARKERS 31
5.2.1 Single microbial source tracking types 31
5.2.2 Samples with mixed microbial source tracking markers 33
6 ENVIRONMENTAL DNA ANALYSIS 35
6.1 MAMMALS 35
6.2 WILD BIRDS 37
7 PATHOGENS 41
7.1 INTRODUCTION TO MULTILOCUS SEQUENCE TYPING 41
7.2 CAMPYLOBACTER 42
7.2.1 Isolation of Campylobacter by culture 43
7.2.2 *Campylobacter* qPCR analysis 44
7.2.3 *Campylobacter* whole genome sequencing 45
7.2.4 *Campylobacter jejuni* whole genome sequencing types 46
7.2.5 *Campylobacter coli* whole genome sequencing types 51
7.3 SALMONELLA 53

7.3.1 Isolation of Salmonella by culture 53
7.3.2 *Salmonella* qPCR analysis 53
7.3.3 Whole genome sequencing of *Salmonella* 54
7.4 SHIGA TOXIN-PRODUCING *E. COLI* 58
7.4.1 Isolation of shiga toxin-producing *E. coli* by culture 58
7.4.2 Shiga toxin-producing *E. coli* qPCR analysis 59
7.4.3 Whole genome sequencing of shiga toxin-producing *E. coli* 60
7.5 PROTOZOA 61
7.5.1 *Cryptosporidium* and *Giardia* detection 62
7.6 ENTERIC VIRUSES 64
7.6.1 Adenovirus 64
7.6.2 Enterovirus 64
7.6.3 Norovirus 64
7.6.4 Virus analysis 64
7.6.5 Virus testing in phases 2 and 3 66
8 PREDICTING THE PRESENCE OF PATHOGENS 68
8.1 INTRODUCTION 68
8.2 ANALYSIS APPROACH 69
8.3 PRESENCE OF *CAMPYLOBACTER* 72
8.4 PRESENCE OF *SALMONELLA* 74
8.5 PRESENCE OF STEC 75
8.6 PRESENCE OF *CRYPTOSPORIDIUM* AND *GIARDIA* 77
8.7 C. JEJUNI WHOLE GENOME SEQUENCING AND MICROBIAL SOURCE TRACKING 79
8.8 FIELD VARIABLES AND BACTERIAL PATHOGEN PRESENCE 79
9 PATHOGEN CONCENTRATIONS 83
9.1 *CAMPYLOBACTER* AND *E. COLI* 83
9.1.1 All data 83
9.1.2 Bathing Season 87
9.2 *CAMPYLOBACTER* AND ENTEROCOCCI 88
9.3 CRYPTOSPORIDIUM AND GIARDIA 90
9.3.1 All data 90
9.4 FIELD VARIABLES AND MST AS PREDICTORS OF CAMPYLOBACTER CONCENTRATION 92
9.4.1 Modelling Approach 92
9.4.2 Results 92
10 INTRODUCTION TO THE QUANTITATIVE MICROBIAL RISK ASSESSMENT simulation model 95
10.1 QUANTITATIVE MICROBIAL RISK ASSESSMENT FRAMEWORK 95
10.2 PROBLEM FORMULATION 95
11 EXPOSURE ASSESSMENT 97
11.1 EXPOSURE PATHWAY 97

11.1.1 Assumption 98
11.2 PATHOGENS AT FRESHWATER SITES 98
11.3 ESTABLISHING THE RELATIONSHIP BETWEEN *CAMPYLOBACTER* AND FIB CONCENTRATIONS 98
11.3.1 Step one – MPN uncertainty 98
11.3.2 Step two – Estimating the *Campylobacter* distribution for *E. coli* concentrations 99
11.3.3 Step two – Estimating the *Campylobacter* distribution for enterococci concentrations 101
11.3.4 Step three – Sampling from the *Campylobacter* distribution 101
11.4 ESTABLISHING THE RELATIONSHIP BETWEEN PROTOZOA AND ENTEROCOCCI 102
11.5 RECREATIONAL ACTIVITIES CONSIDERED 104
11.6 POPULATION GROUP CONSIDERED 105
11.7 HOW MUCH WATER IS LIKELY TO BE INGESTED? 105
11.7.1 Ingestion rate 105
11.7.2 Duration of swimming activity 106
11.8 SIMULATED VOLUME OF WATER INGESTED 106
11.8.1 Assumptions 107
12 HEALTH EFFECTS ASSESSMENT 109
12.1 INTRODUCTION 109
12.2 HEALTH CONSEQUENCES 110
12.3 DOSE-RESPONSE 110
13 RISK CHARACTERISATION 113
13.1 SIMULATION APPROACH 113
13.2 CAMPYLOBACTERIOSIS RISK 114
13.2.1 Campylobacteriosis by *E. coli* concentration 114
13.2.2 Campylobacteriosis by enterococci concentration 117
13.2.3 Sensitivity and uncertainty analyses 120
13.3 CRYPTOSPORIDIUM AND GIARDIA RISK 121
13.3.1 Simulation results 121
13.3.2 Uncertainty in simulation estimates 121
13.4 RISK CHARACTERISATION CONCLUSIONS 123
14 ESTIMATING RISK OF NOROVIRUS INFECTION ON THE BASIS OF CRASSPHAGE LEVELS 124
14.1 INTRODUCTION 124
14.2 USE OF CRASSPHAGE LEVELS TO ESTIMATE THE RELATIVE LEVEL OF NOROVIRUS 125
14.2.1 Estimating the norovirus concentration in water at swimming site 126
14.2.2 Simulation model results 126
14.3 *E. COLI* AS A SCREENING TOOL FOR CRASSPHAGE 128
15 COMPARISONS TO THE 2002 FRESHWATER PROGRAMME 130

15.1 DATA COLLECTION AND LABORATORY ANALYSIS 130
15.2 COMPARISON OF THE 2002 AND 2025 CAMPYLOBACTERIOSIS QMRA MODELLING APPROACHES 135
15.2.1 Model components 135
15.2.2 Model results 138
16 DISCUSSION AND CONCLUSIONS 139
16.1 CAMPYLOBACTER 139
16.2 OTHER PATHOGENS 141
16.3 MICROBIAL SOURCE TRACKING 141
16.4 QMRA APPROACH TO SUPPORT NATIONAL LEVEL METRICS 142
16.4.1 Single verses multiple pathogen QMRA estimates of public health risk 143
16.4.2 Water ingestion 144
16.4.3 Pathogen and indicator concentrations 145
16.4.4 Dose-response 145
16.5 E. COLI AND ENTEROCOCCI AS HUMAN HEALTH RISK INDICATORS ... 146
16.6 SITE SPECIFIC QMRA 149
16.7 CONCLUSIONS 149
APPENDIX A: HEALTH EFFECTS 151
A.1 CAMPYLOBACTERIOSIS 151
A.1.1 Sequelae 151
A.1.2 Burden of disease 153
A.2 SALMONELLOSIS 154
A.2.3 Sequelae 154
A.2.4 Burden of disease 155
A.3 SHIGA TOXIN-PRODUCING *E. COLI* INFECTION 155
A.3.1 Sequelae 155
A.3.2 Burden of disease 156
A.4 CRYPTOSPORIDIOSIS 157
A.4.1 Sequelae 157
A.4.2 Burden of disease 158
A.5 GIARDIASIS 158
A.5.1 Sequelae 158
A.5.2 Burden of disease 158
A.6 NOROVIRUS INFECTION 159
A.6.1 Burden of disease 159
APPENDIX B: METHODS 160
B.1 CULTURE METHODS AND IDENTIFICATION OF BACTERIA BY PCR 160
B.1.1 Total coliforms, *E. coli* and enterococci 160
B.1.2 Campylobacter 160
B.1.3 *Salmonella* 161
B.1.4 Shiga toxin-producing *E. coli* 162
B.2 MICROSCOPY METHODS 162

B.3 WHOLE GENOME SEQUENCING 163
B.4 QUANTITATIVE PCR 163
B.4.5 Microbial source tracking markers, bacteria and protozoa 163
B.4.6 eDNA Analysis 168
B.4.7 Viruses 168
B.5 NUMERICAL ANALYSES 168
B.5.1 Ingestion rate distribution fitting 168
B.5.2 Quantile Regression 169
B.5.3 Regression modelling 169
APPENDIX C: SAMPLING SITES 170
APPENDIX D: ENVIRONMENTAL DATA 174
D.1 FIELD DATA COLLECTION 174
D.2 CONDUCTIVITY, PH AND DISSOLVED OXYGEN 175
D.3 WATER TEMPERATURE AND SEASON 176
APPENDIX E: DOSE RESPONSE 177
E.1 CAMPYLOBACTERIOSIS 177
E.2 CRYPTOSPORIDIOSIS 178
E.3 GIARDIASIS 180
APPENDIX F: SENSITIVITY ANALYSIS 182
F.1 INTRODUCTION 182
F.2 SIMULATION ITERATIONS AND REPORTING ACCURACY 182
F.3 WATER INGESTION RATE 182
F.3.1 *Variable verses fixed ingested water volumes* 182
F.3.2 *Independence of water ingestion rate and duration of swimming activity* 183
F.4 ANALYSIS BIAS IN COLLECTION OF PROTOZOA DATA IN LOW VOLUMES OF WATER 185
APPENDIX G: GLOSSARY 186
APPENDIX H: ABBREVIATIONS 191

LIST OF TABLES

Table 1 Estimated percentage of swims in water with different *E. coli* and enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims.. 5

Table 2: 2023-2024 New Zealand notifiable disease rates for relevant gastrointestinal infections..... 17

Table 3: Number of samples collected across the different observed land use categories .. 21

Table 4: Faecal indicator bacteria concentrations by culture results..... 23

Table 5: *E. coli* concentrations in collected freshwater samples 24

Table 6: *E. coli* based gradings of rivers sampled at least 10 times during the study for all data..... 25

Table 7: qPCR concentration estimates for *E. coli* and enterococci..... 26

Table 8. Comparison of *E. coli* qPCR and MPN culture quartiles. 27

Table 9. Comparison of Enterococci qPCR and MPN quartiles. 27

Table 10: Microbial source tracking marker prevalence and gene copy (GC)/100 mL concentration in the collected water samples 29

Table 11: Concentration ranges for categorising MST makers. 32

Table 12: Categorisation of water samples by microbial source tracking marker subgroups 32

Table 13: Microbial source tracking marker groupings 34

Table 14: Animal eDNA detected 36

Table 15: Comparison of eDNA from cattle and sheep with Ruminant BacR MST marker .. 37

Table 16: Levels of bird eDNA detected 38

Table 17: Regional distribution of wild bird eDNA..... 39

Table 18. Detection of MST GFD marker compared with eDNA detection of wild bird..... 40

Table 19 Detection of MST GFD marker compared with detection of eDNA from Mallard Duck. 40

Table 20: *Campylobacter* prevalence and concentrations for all collected freshwater samples 43

Table 21: Frequency of isolation of *C. jejuni* and *C. coli* at different sampling locations 44

Table 22: Comparison of MPN-based culture analysis with qPCR for detection of *C. jejuni* 44

Table 23: Comparison of MPN-based culture analysis with qPCR for detection of *C. coli* ... 45

Table 24: Other species of *Campylobacter* isolated from water samples..... 45

Table 25: Frequency of isolation of *Salmonella* from water samples 53

Table 26: Comparison of MPN-based culture analysis and qPCR for detection of *Salmonella* 54

Table 27: Frequency of serotypes of sequenced *Salmonella* in water samples. 55

Table 28: Frequency of detection of STEC genes from broths and pooled isolates 58
Table 29: Frequency of isolation of STEC in broths from different sites..... 59
Table 30: Summary of results from sequencing of putative STEC isolates 60
Table 31: Summary of *Cryptosporidium* and *Giardia* concentrations in water samples..... 62
Table 32: Frequency of isolation of *Cryptosporidium* and *Giardia* at different sites..... 63
Table 33: Viral qPCR results for samples tested in Phase I..... 65
Table 34: Odds ratios for the presence of *C. jejuni* and or *C. coli* in water samples as predicted by log_{10} scaled indicators using the univariate GLMM model. 73
Table 35: *C. jejuni* isolation from freshwater samples with different MST profiles. Ranked from highest to lowest detection of *C. jejuni*. 73
Table 36: Odds ratios for the presence of *Salmonella* in water samples as predicted by log_{10} scaled indicators using univariate GLMM. 74
Table 37: *Salmonella* presence by culture in samples with different MST profiles. 74
Table 38: Odds ratios for the presence of STEC in water samples as predicted by log_{10} scaled indicators using univariate GLMM. 76
Table 39: Samples with shiga toxin genes detected in broth culture compared with MST source. 76
Table 40: Odds ratios for the presence of *Cryptosporidium* in water samples as predicted by log_{10} scaled indicators using univariate GLMM. 77
Table 41: Odds ratios for the presence of *Giardia* in water samples as predicted by log_{10} scaled indicators using univariate GLMM. 77
Table 42: Samples with *Cryptosporidium* or *Giardia* detected compared with MST source. 78
Table 43: Odds ratios for the presence of bacterial pathogens in water samples as predicted by environmental field variables using a univariate GLMM model. 81
Table 44: Number of samples, sites and detections of *C. jejuni* and/or *C. coli* in each *E. coli* concentration interval. 85
Table 45: Health effects of infection with pathogens considered in this study and norovirus 111
Table 46: Dose response relationships used in the Monte Carlo simulation model considering variability in response to ingestion of the pathogen 112
Table 47: Estimated percentage of swims in water with different *E. coli* concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. 115
Table 48: Estimated percentage of swims in water with different enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. 118
Table 49: Estimated percentage of swimming in water with different enterococci concentrations resulting in *Cryptosporidium* or *Giardia* infection, illness and DALYs per 1000 swims..... 121

Table 50 Risk of infection calculated from copies of CrAssphage, and number of samples in each category (excluding samples identified by MST as including aged human faecal contamination). 127

Table 51. Comparison of key sampling and laboratory analytical differences between the 2002 and 2025 studies. 132

Table 52: High level comparison of the QMRA simulation models reported in 2002 and 2025. 136

Table 53 Comparison of the percentage of swims resulting in *Campylobacter* infection (2002 model) and infection or illness (2025 model) for swimming in water with given values of *E. coli* concentrations. 138

Table 54 Estimated percentage of swims in water with different *E. coli* and enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. Based on all data collected. 147

Table 55: Gastroenteritis rates observed in freshwater swimming trials in water with different *E. coli* or enterococci concentrations adapted from Wiedenmann et al (2006). 148

Table 56: Bacteria and protozoa qPCR target organism or gene, primers and probes, and assay concentrations and thermocycling conditions for bacteria, protozoa and MST markers. 165

Table 57: qPCR target organism or gene, assay concentrations and specific amplifying conditions and methods for SYBR Green assays of MST markers. 167

Table 58: Location of sampling sites and observed dominant contamination sources 170

Table 59: Physiochemical parameters measured in water samples 174

Table 63: Comparison of simulation outputs for three different campylobacteriosis dose-response models, using data collected over the full year. 178

Table 64: Comparison of simulation outputs using enterococci as the FIB for two different cryptosporidiosis dose-response models. 180

Table 65: Comparison of simulation outputs using enterococci as the FIB for two different giardiasis dose-response models. 181

Table 66: Simulation model estimation of Campylobacter infection given swimmers ingesting selected volumes of water (rounded to 0.5%). 183

Table 67: Simulation model estimation of campylobacteriosis infection given swimmers ingesting volumes of water with different dependencies between the water ingestion rate and duration of the swim event. 184

Table 68: Comparison of simulation outputs using enterococci as the FIB for *Cryptosporidium* infection and illness with and without imputation of missing *Cryptosporidium* concentrations at low volumes of filtered water. 185

LIST OF FIGURES

Figure 1: Study timeline 12

Figure 2: Site locations and number of samples collected during the three phases of this study 20

Figure 3: Concentrations of faecal indicator bacteria in collected freshwater samples..... 23

Figure 4: *E. coli* and enterococci concentrations by culture and qPCR..... 26

Figure 5: Box and whisker plots for microbial source tracking marker concentrations in the collected water samples 30

Figure 6: Whole genome MLST minimum spanning tree of *C. jejuni* ST2381 isolates 47

Figure 7: Whole genome MLST minimum spanning tree of *C. jejuni* ST45 isolates 48

Figure 8: Whole genome MLST minimum spanning tree of *C. jejuni* ST3655 isolates 49

Figure 9: Whole genome MLST minimum spanning tree of *C. jejuni* isolates with STs observed 5 – 19 times 50

Figure 10: Whole genome MLST minimum spanning tree of *C. coli* isolates 52

Figure 11: Whole genome MLST minimum spanning tree of *Salmonella* Typhimurium 56

Figure 12: Whole genome MLST minimum spanning tree of non-Typhimurium *Salmonella* serotypes isolated more than once Taranaki 57

Figure 13: Whole genome MLST analysis of STEC isolates..... 61

Figure 14: Distribution of *Cryptosporidium* and *Giardia* concentrations in water samples 62

Figure 15: The presence/absence and probability of detection of STEC by GFD concentration. 70

Figure 16: The presence/absence and probability of detection of *Campylobacter* by GFD concentration. 71

Figure 17: Comparison of *Campylobacter* (combined *C. jejuni and C. coli*) and *E. coli* concentrations in freshwater samples. 84

Figure 18: *Campylobacter* concentration and interquartile range (25th to 75th percentile) ... 86

Figure 19: Comparison of *Campylobacter (C. jejuni and C. coli)* and *E. coli* concentrations in freshwater samples during (circles) and outside (triangles) the bathing season. 87

Figure 20: Comparison of *Campylobacter (C. jejuni and C. coli)* and enterococci concentrations in freshwater samples. 88

Figure 21: Comparison of *Campylobacter (C. jejuni and C. coli)* concentrations with enterococci concentrations in freshwater samples during (circles) and outside (triangles) the bathing season. 89

Figure 22: Comparison of *Cryptosporidium* and *Giardia* concentrations with *E. coli* or enterococci concentrations in freshwater samples. 90

Figure 23: Comparison of *Cryptosporidium and Giardia concentrations compared to E. coli or* enterococci concentrations in freshwater samples during (circles) and outside (triangles) the bathing season..... 91

Figure 24: Quantitative microbial risk assessment framework. Adapted from World Health Organization (2016). 96

Figure 25: Exposure pathway for single freshwater swimming event..... 97

Figure 26: Quantile regression estimates for *Campylobacter* distributions with different *E. coli* concentrations 99

Figure 27: Quantile regression estimates for *Campylobacter* distributions with *E. coli* concentrations using (a) all data and (b) bathing season data only 100

Figure 28: Quantile regression estimates for *Campylobacter* distributions with enterococci concentrations using (a) all data and (b) bathing season data..... 101

Figure 29: Quantile regression estimates for *Cryptosporidium* and *Giardia* distributions with different enterococci concentrations..... 103

Figure 30: Model distributions for the volume of water ingested by 6 - 12-year-olds during open-air swimming..... 108

Figure 31: Health effects flow diagram 109

Figure 32: Monte Carlo simulation model..... 113

Figure 33: Estimated *Campylobacter* illness occurrence and 50th, 75th and 95th percentiles of individual probability of illness and *Campylobacter* dose in water with different *E. coli* concentrations..... 117

Figure 34: Estimated *Campylobacter* illness occurrence and 50th, 75th and 95th percentiles of individual probability of illness and *Campylobacter* dose in water with different enterococci concentrations. Data collected all year round or collected during the bathing season. 119

Figure 35: Linear correlation (dashed blue line) between CrAssphage concentration in wastewater influent and: (a) norovirus genogroup GI (NoV GI); (b) norovirus genogroup GII (NoV GII); (c) enterovirus (EV); and (d) human adenovirus (AdV) concentration. 125

Figure 36 Scatterplot of CrAssphage concentration (GC/100ml) compared with *E. coli* (MPN/100ml)..... 128

Figure 37: Flow diagram showing the structure of the 2002 and 2025 simulation mode ... 137

Figure 38: An example of a *Campylobacter* PCR detection using the MultiNA electrophoretic system. 161

Figure 39: Dissolved oxygen data recorded during the three phases of the study, coloured by North or South Island with time phased weekly medians by island. 175

Figure 40: Water temperature data from the three phases of the study, coloured by North and South Island sites and showing time phased weekly medians by island. 176

Figure 41: Exponential (solid line) and Fractional Poisson dose-response models for cryptosporidiosis. 179

Figure 42: Simulation swim event duration and water ingestion rate. 184

EXECUTIVE SUMMARY

Background

Freshwater recreation has an important role in the health and wellbeing of New Zealanders. There is also a significant economic impact of freshwater recreation, both on domestic and international tourism, and on the management steps taken to progress water quality improvements. Internationally, most guidelines have been derived based on epidemiological studies undertaken where waters were largely impacted by human sewage. This approach is not feasible in New Zealand due to (i) the low population levels and number of people attending each freshwater site, and (ii) the rural nature of many freshwater catchments close to recreational sites, that is likely to lead to wild or livestock animal faecal contamination, as well as human.

A Quantitative Microbial Risk Assessment (QMRA) is an alternative modelling approach which can be used to support the derivation of water quality guidelines. QMRA uses mathematical models with quantitative data, such as the concentrations of pathogens in the water and the volume of water ingested during recreation, to estimate rates of infection or illness.

A QMRA was undertaken in 2002 using freshwater field data generated in New Zealand between 1998 to 2000 (Freshwater Microbiology Research Programme (FMRP) McBride et al, 2002). The results from this analysis were used to inform the 2003 Ministry for the Environment and Ministry of Health Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas (hereafter, the Guidelines). These guidelines were developed to help water managers monitor, report on and control the public health risk posed by microbiological contamination of recreational waters (Ministry for the Environment, 2003).

In the last 20 years there have been significant developments in land management practices and analytical technologies, creating a recognised need to repeat both the survey of rivers and the QMRA analysis.

In 2017, the Ministry for the Environment commissioned a work to identify appropriate rivers for sampling, to review microbial methods, QMRA components and to design a survey and QMRA. This included a stakeholder workshop held in Wellington in May 2018. A pilot project was commissioned in 2020, with sample collection, analysis and QMRA completed using two subsequent phases. This report describes analysis of collated data from the pilot project and the two subsequent phases.

Freshwater survey

Between February 2020 and June 2024, 1,041 samples were collected from 71 rivers across New Zealand. Samples were collected in three phases – February to March 2020 (52 samples), January to May 2021 (256 samples), and September 2022 to June 2024 (733 samples). Sites were predominantly chosen with known potential faecal inputs, so that the relationships between indicator variables and pathogens could be explored. While the sites cover a wide range of locations across New Zealand, the key findings below are not an

indication of the state of rivers in New Zealand. Instead, they show that sufficient data was collected, at low, medium and high levels of indicators and pathogens, to facilitate an understanding of the relationships between indicators and pathogens.

Key findings from analysis of these samples were:

- A third of samples had less than 130 *Escherichia coli* per 100 ml, and a third more than 550 *E. coli* per 100 ml providing a useful range of contamination levels.
- Microbial source tracking markers for human faecal material were identified in 35% of samples, ruminant faecal sources in 66% of samples, and wild bird faecal markers in 80% of samples. Seventy percent of samples had more than one faecal source.
- eDNA analysis indicated contamination from 19 different mammals (cattle, rat, possum, sheep, and deer the most frequently observed), and 40 different wild birds (mallard duck and pūkeko the most frequently observed), highlighting the need to consider a wide range of potential sources of faecal contamination.
- *Campylobacter* were isolated from 68% of samples, with 4% of samples having greater than 100 *C. jejuni* per 100 ml, 19% between 10 and 100 *C. jejuni* per 100 ml, and 19% between 1 and 10 *C. jejuni* per 100 ml. *C. coli* was detected in 13% of samples, but only 3% of samples had greater than 1 *C. coli* per 100 ml.
- *Cryptosporidium* and *Giardia* were detected in 50% and 69% of samples respectively, but only 1% of samples had greater than 10 *Cryptosporidium* per 10 litre and 3% of samples had greater than 10 *Giardia* per 10 litre.
- *Salmonella* were detected in up to 19% of samples, and shiga toxin-producing *E. coli* (STEC) *stx1* or *stx2* genes in 20 – 30% of broths generated from samples of one litre. Levels of each organism were almost always <1 per 100 ml, with only 1% of samples yielding a whole genome sequenced STEC.
- Pathogenic viruses were detected at very low levels in the first phase of the study, therefore, a non-pathogenic indicator virus of human faecal contamination (CrAssphage) was evaluated as a surrogate to assess viral risk.

Data analysis

The relationship between pathogens (*Campylobacter* (*C. jejuni* and *C. coli*), *Salmonella*, STEC, *Cryptosporidium*, and *Giardia*), faecal indicator bacteria (*E. coli*, enterococci), Microbial Source Tracking (MST) markers (human faecal: CrAssphage, HF183, BiADO; ruminant: BacR and avian: GFD) and biophysical parameters were explored, with the following relationships identified.

- Generalised linear mixed model analysis indicated that increasing *E. coli* and enterococci were positively associated with increases in the detection of *Campylobacter*, *Salmonella* and STEC, but not *Cryptosporidium* or *Giardia*.
- *E. coli* and enterococci are a weaker predictor of *Cryptosporidium* and *Giardia* concentrations, which is likely due to the very low concentrations of *Cryptosporidium* and *Giardia* found in the survey.
- The concentration distribution of *C. jejuni* and *C. coli* shows an increasing trend with increasing *E. coli* concentration up to approximately 1,000 *E. coli* MPN/100 ml, at which point there is little change to the *Campylobacter* concentration distribution.

- A similar trend was observed for enterococci, with increasing trend in *Campylobacter* concentrations with increasing enterococci concentrations.
- *C. jejuni* were associated with human, ruminant and wild bird sources, as defined by MST. Linear mixed regression modelling found a relationship of increasing *Campylobacter* concentration with increasing wild bird and ruminant MST markers. A higher prevalence and higher concentrations of *Campylobacter* were associated with samples with multiple faecal sources rather than a single source.
- *Salmonella* and *Giardia* were potentially more associated with human sources, STEC with ruminant sources, and *Cryptosporidium* prevalence did not differ significantly in samples with different MST marker sources.
- There was evidence of decreasing detection of *Campylobacter* with increasing water temperature, dissolved oxygen or pH, and increasing detection of *Campylobacter* with increasing turbidity, or increasing rainfall in the one to three days before water sampling.
- The non-bathing season has higher odds of *Campylobacter* detection than in the bathing season, as well as higher *Campylobacter* concentrations in water when there are lower concentrations of *E. coli.* This difference between the two periods is probably due to lower survival of *Campylobacter* in higher water temperatures (and increased sunlight) and translates to a slightly lower risk estimate for campylobacteriosis in the bathing season (31st October to 31st March), than when using the whole dataset.

QMRA Analysis

QMRA for recreational water activities is a formal quantitative risk assessment tool, that combines knowledge and data about the presence of pathogens at freshwater sites, with human exposure pathways and possible health effects from being exposed to the pathogens.

Monte Carlo simulation models are an established QMRA tool which allow the replacement of point estimates of model components with variability distributions, replicating the variability seen in the real world. The simulation model estimates the illness and/or infection risk for swimmers at freshwater sites for given indicator bacteria, accounting for the variability of (i) swimmer behaviour and (ii) the concentrations of indicators and pathogens in the water as identified from the survey data. Simulation models have been developed to estimate the risk of infection or illness for a person swimming at a freshwater site with water with a given concentration of *E. coli* or enterococci.

The advantage of the dataset generated by this study is that no assumptions need to be made about the fate or survival of the pathogens in the water before a swimmer ingests the water - the water samples provide the indicator, pathogen, MST marker information for the same sample of water.

Due to the large dataset, quantile regression analysis could be used to generate *Campylobacter*, *Cryptosporidium* and *Giardia* concentration distributions for a given *E. coli* or enterococci concentration. For a given indicator concentration (*E. coli* or enterococci), each iteration of the simulation model takes samples from the pathogen concentration distribution and combines it with a volume of water ingested to give the dose of a pathogen for a single

swimming event. For this simulation model the volume of water ingested is based on the behaviour of children aged 6- to 12-years-old, as these have the highest water ingestion rates while swimming and tend to swim for longer than adults. The distribution of volume of water ingested defines 50% of people will ingest 44 ml or less (equivalent to three tablespoons), and only five percent of swimmers would ingest more than 270 ml of water.

Once the dose for the swimmer is determined, dose-response relationships are applied to calculate the probability a swimmer will become infected given the dose ingested and the probability the person will become ill if infected. The simulations are repeated multiple times and the outcomes of infection and illness recorded for each swimmer, allowing infection and illness rates to be calculated.

Based on the national freshwater data collected in this study, the simulation model demonstrated that *Campylobacter* results in a higher percentage of swimmers becoming ill from swimming in water with given *E. coli* or enterococci concentrations, than *Cryptosporidium* or *Giardia*.

The simulation modelling has also shown that the concentrations of the indicators *E. coli* or enterococci in water can be used to define changes in the percentage of swims resulting in *Campylobacter* infection or illness, and these are summarised in Table 1 for a range of indicator values. Disability adjusted life years (DALY) per 1,000 swims is an additional metric which incorporates the illness burden, with 1 DALY equal to a loss of one year of quality life either through years of life lost or due to living with illness or disability.

Table 1: Estimated percentage of swims in water with different *E. coli* and enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. Based on all data collected.

MPN / 100 ml	Log_{10} /100 ml	Percentage of swims resulting in *Campylobacter* infection		Percentage of swims resulting in illness (campylobacteriosis)		DALY due to campylobacteriosis per 1,000 swims	
		E. coli	Enterococci	*E. coli*	Enterococci	*E. coli*	Enterococci
40	1.6		4.5		2.0		0.3
50	1.7	2.0		1.0		0.1	
100	2.0	3.0	6.0	1.0	2.5	0.2	0.4
200	2.3	4.5	7.5	2.0	3.0	0.3	0.5
300	2.5	6.0	8.5	2.5	3.5	0.4	0.5
400	2.6	7.0	9.0	3.0	3.5	0.4	0.5
500	2.7	8.0	9.5	3.0	4.0	0.5	0.6
600	2.8	9.0	10.5	3.5	4.0	0.5	0.6
800	2.9	9.5	12.0	4.0	5.0	0.6	0.7
1,000	3.0	9.5	12.5	4.0	5.0	0.6	0.8
130	2.1	[0.0][a] 3.5		1.5		0.2	
260	2.4	[0.9] 5.5		2.0		0.3	
540	2.7	[4.9] 8.5		3.5		0.5	

Notes:

a: [ ] Percentage of swims resulting in *Campylobacter* infection as estimated by the 2002 model (McBride et al., 2002).

In consideration of the modelling uncertainties, the percentage of swims estimated to result in infection are rounded to nearest multiple of 0.5%, and DALY estimates are rounded to one decimal place.

Key insights from this simulation modelling include:

- Even at low *E. coli* (50 MPN/100 ml) or enterococci (40 MPN/100ml) concentrations, there can be a low risk of campylobacteriosis (1-2% of swims resulting in illness).
- At an *E. coli* concentration of 260/100 ml, the model estimates a 2.0% risk of illness, while at 540 *E. coli*/100 ml this risk increases to 3.5%. These are concentrations of *E. coli* used in the current Guidelines. However, it should be noted that the current guidelines are linked to infection rather than illness outcomes.
- The relationship between the concentrations of *E. coli* and *Campylobacter* is slightly different when considering data collected during the bathing season, compared to data collected all year round. The difference is sufficient to lower the percentage of swims resulting in infection by between 0.5 and 1.5% using data only from the bathing season instead of using data collected all through the year.
- The previous QMRA study (McBride et al, 2002), reported infections, rather than illness, with the estimated percentage of swims resulting in infections lower than in

this study. However, the modelling approach has been updated, making the two models not directly comparable. The previous modelling approach also allowed a zero risk to be determined (Table 1). While pragmatic, this masks the reality that zero risk is not achievable, and there will always be a risk of illness in water with possible faecal inputs.

Human faecal pollution of rivers has additional risks of infection from human infective viruses. Modelling based on the human MST marker CrAssphage suggests illness risks from norovirus of >3% when 30,000 genome copies/100 ml of CrAssphage detected. Eight per cent of the samples tested during the survey exceeded this level. These were from 8 different rivers, all with elevated *E. coli* concentrations. This risk is in addition to the risk identified in the *Campylobacter*-based QMRA and emphasises the importance of human sewage as a risk to human health. In samples with greater than 1,000 *E. coli*/100 ml, the presence of high CrAssphage concentrations may increase the risk of illness from 4% when considering just campylobacteriosis to 10% or greater considering campylobacteriosis and norovirus. Other pathogens will also create a cumulative risk in the most contaminated samples, while in other samples they may be the only risk.

Comparisons of New Zealand Recreational Water Guidelines with overseas approaches have been undertaken previously (McBride and Soller, 2017, Leonard and Eaton, 2021, Leonard, 2025). These reviews included guidelines from WHO (2021), Australia (2008), EU (2006), US EPA (2012) and Canada (2012). In this study, illness as the endpoint for health outcomes aligns more closely with international approaches. While there is variation in the statistical metrics used in guidelines, the findings in this study do align with international values.

Key limitations and uncertainties

The simulation modelling approach randomly samples from distributions to incorporate real life variability in the combinations of ingestion of pathogens and health response. Outputs were found to be stable to within 0.22% or 2.2 illnesses per 1,000 swim events for 50,000 or more iterations, which is the number of iterations used in the model. Infection and illnesses are reported in 0.5% intervals to reflect this uncertainty.

Regional differences in the duration of swimming events are not included in the modelling as there is no data available to quantify the magnitude of any differences. While the coldness of water may affect swimming duration, tiredness and other physical restrictions may reduce differences, particularly on warmer days across the country.

One of the largest uncertainties is the number of organisms required to cause illness or infection which will be variable for different people. The dose response values used in this QMRA (and all other QMRAs) are based on a limited number of studies, mainly based on adult feeding or vaccine trials, with limited low dose data. Immunocompromised individuals will have a higher risk of illness than indicated by the modelling. There is insufficient data available to assess whether this existing dose-response data under or overestimates actual dose response of the 6-12 year old age grouping on which the QMRA was based. The 6-12 year old age group was chosen as a conservative approach to risk assessment, as they are likely to ingest more water during a visit to a freshwater site than other age groups.

There is also uncertainty in the distribution of water volume ingested during swimming, with only overseas generated data available. The model outputs, however, are less sensitive to ingestion volume than the choice of dose-response model.

The water samples collected in this study, were from undisturbed water 200 mm below the surface. Sediments may be a significant source of pathogens, including protozoa. Inclusion of stirred sediment samples may increase the pathogen loadings, and therefore, probability of infection and illness.

Conclusions

This study provides an improved and expanded data set describing the relationships in New Zealand rivers between a range of microbial indicators and pathogens and an updated simulation model for estimating the risk to swimmers for gastrointestinal pathogens. Consistent with the previous FMRP study, the risk of infection following exposure to river water can be estimated based on concentrations of *E. coli* and the risk of *Campylobacter* infection. The percentage of swims resulting in infection, is predicted to be slightly higher than predicted in the 2002 model, for a given *E. coli* concentration. This current modelling extends the analysis to risk of illness which is more in line with international approaches.

Both *E. coli* and enterococci are useful indicators of faecal contamination and of potential health risk. Increasing concentrations of these faecal indicators are associated with increasing risk from faecally-associated pathogens.

A risk of illness is present across all rivers, and from all sources of contamination. Wild birds contribute faecal material to most rivers and create a background risk of illness in addition to any anthropogenic inputs. Anthropogenic sources do, however, increase that risk. Ruminant faecal sources increase the likelihood of *Campylobacter* and STEC detection, while human faecal sources increase the likelihood of *Campylobacter* and *Salmonella* detection. Human faecal matter is also the source of human infective viruses which may need to be considered when evaluating a site.

Elevated levels of faecal indicators should prompt the use of MST tools to understand potential sources of contamination. In the case of human sources being identified, the concentrations of the MST marker CrAssphage can be used to estimate potential norovirus risk. MST also provides guidance to water managers for mitigation opportunities which can support water quality improvements.

Recreational water guidelines are a science-informed policy decision. Analytical methodologies, pathogen carriage, land management practices, and QMRA modelling approaches have all changed since the FMRP study carried out over 20 years ago, and those changes are reflected in the conclusions of this current study. Revision of New Zealand's recreational water quality guidelines also needs to include revisiting sampling, statistical analysis, responses to elevated detections, and alignment with international guidelines. The data and QMRA modelling generated in this study allows for implications of any adjustments to recreational guidelines to be assessed against health risks, using science that is both current, and New Zealand specific.

1 INTRODUCTION

1.1 BACKGROUND

Pathogenic microorganisms are bacteria, viruses and protozoa that are capable of causing illness or disease in humans. The presence of pathogens in recreational waters poses a health risk to those using the water for swimming, food gathering, and other contact activities. Most waterborne pathogens that cause human illness, including *Campylobacter* spp., *Salmonella* spp., norovirus, *Giardia* spp. and *Cryptosporidium* spp., are associated with human and/or animal faeces. The contamination of waterways with faecal material may result from the discharge of inadequately treated sewage, leaking sewage pipes or septic tanks, run-off from urban and/or agricultural land, and direct deposition from farm or wild animals. Health effects resulting from contact with contaminated water include gastrointestinal illness, respiratory illness and skin infections.

In New Zealand, the microbial quality of recreational waters is monitored by regional and unitary councils (hereafter referred to as Regional Councils). Direct monitoring for the presence of pathogens in water is impractical, as pathogens tend to be present intermittently and at low concentrations, and analyses are complex and expensive. Rather, 'indicator organisms' are used to monitor microbial water quality. Indicator organisms are not usually pathogenic, however, as they are typically found in the intestinal tract of warm-blooded animals, they are indicative of faecal contamination and the potential presence of pathogens. The most commonly used indicators of faecal contamination are *Escherichia coli* (*E. coli*), faecal coliforms and enterococci, the presence of which are quick and cost-efficient to enumerate compared with pathogens.

The 2003 Ministry for the Environment and Ministry of Health Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas (hereafter 'the Guidelines') (Ministry for the Environment, 2003) were developed to help water managers monitor, report on and control the public health risk posed by microbiological contamination of recreational waters. They include information on undertaking catchment and site assessments for possible contamination sources, and methods and standards for measuring and monitoring microbial water quality.

The freshwater component of the Guidelines uses the indicator organism *E. coli* to assess water quality, with numeric guideline values developed from the findings of the 1998-2000 Freshwater Microbiology Research Programme (FMRP) (McBride et al, 2002, Till et al. 2008). The FMRP included a nationwide survey of microbial water quality, monitoring 25 sites representing different land uses and associated faecal impacts (dairy farming, beef and sheep farming, municipal, wild birds and forested/undeveloped), with water samples collected fortnightly for 15 months, and analysed for 10 pathogens and indicators. Using these data, a Quantitative Microbial Risk Assessment (QMRA) for campylobacteriosis was performed as *Campylobacter* was the pathogen most frequently detected. A QMRA is a framework to combine information on the particular pathogen(s) and the potential dose or exposure (a function of the concentration of pathogens in the water and the volume of water that might be ingested during recreation), to estimate the risk of infection and illness.

The National Objective Framework (NOF) of the National Policy Statement for Freshwater Management (NPS-FM) 2020 (New Zealand Government, 2020) and its associated amendments direct Regional Councils to set objectives and limits to manage freshwater rivers and lakes for two compulsory national values: ecosystem health, and human health and recreation. The NOF includes *E. coli* as an attribute, with various numeric categories which are based on the Guidelines.

The FMRP report (McBride et al, 2002) was not prescriptive of the content of the Guidelines. Rather, the Guidelines were developed by a working group who used the FMRP results incorporated into a QMRA model. In so doing, the key decisions made by the working group, and authorised by the Ministry for the Environment and the Ministry of Health were that:

1. The freshwater component of the Guidelines should be based on the risk of *Campylobacter* infection (noting that some individuals infected with that bacterium may not become ill) and its (moderate) association with *E. coli* concentrations.

2. The Guidelines should adopt the World Health Organisation (WHO) 'Annapolis Protocol' (World Health Organization, 1999, 2003), in which grading and surveillance are separate, yet complementary, elements.

3. The grading outcome for a recreational site should take account of both the *E. coli* data and a sanitary survey of the site to assess its susceptibility to faecal contamination.

1.2 RATIONALE FOR A NEW SURVEY AND QMRA

An essential underpinning of the Guidelines is a robust understanding of the relationship between the pathogens that may be present in the water, and the indicator organisms used to monitor water quality (Ministry for the Environment, 2003). Since the 2003 Guidelines were developed, there have been significant advances in laboratory methods, including the development and routine use of molecular techniques, such as quantitative polymerase chain reaction (qPCR) and whole genome sequencing (WGS). These methods allow for more sensitive and specific detection of microorganisms, which should allow for improved characterisation of the relationship between the concentrations of various pathogens and concentrations of indicators. Further, these techniques allow for the characterisation of indicators and pathogens based on their DNA, including information on the serotype, genotype and/or potential virulence of pathogens. For example, not all strains of a species may cause illness in humans, or may cause illness of varying severity (e.g., most *E. coli* are harmless, however a small percentage are pathogenic). Molecular analysis can also assist in providing information as to the likely source of contamination (egg, human, cattle, sheep, wild birds), by investigating the presence of particular microorganisms or particular genotypes of widely dispersed microorganisms that are uniquely associated with a certain host animal. Source attribution is important, as there is a growing appreciation of the different health risks that might be posed by faecal contamination from different animal sources. Understanding sources can also inform the development of mitigation strategies.

The increase in computing power and further development of QMRA, dose-response and quantitative modelling approaches since 2000 allows the QMRA approach used in the 1998-2000 project to be updated and improved.

In addition to improved methodologies, there have also been significant changes in land use patterns within New Zealand, as well as in land management practices (e.g., stock exclusion from waterways, irrigation of agricultural effluent to land). These may affect how microorganisms are introduced to and survive in waterways, and hence the relationship between indicators and pathogens in the water.

Since at least 2013 (Bolton-Ritchie et al, 2013), stakeholders, including central and regional government, Dairy NZ, Beef + Lamb NZ, Medical Officers of Health, and researchers have expressed the need to revise and update the science underpinning the Guidelines (Milne et al, 2017; Moriarty et al, 2018).

1.3 STUDY OUTLINE

A timeline for the study is given in Figure 1.

In 2017, the Ministry for the Environment commissioned a Freshwater Microbiological Sciences Review (Stage One) with four workstreams:

- Work Stream A: Site Selection (Milne et al, 2018)
- Work Stream B: Microbial Methods (Gilpin et al, 2018)
- Work Stream C: Dose Response, Illness and Infection Risk (Horn et al, 2018; Lake et al, 2018)
- Work Stream D: Survey Design and QMRA Project Plan (Moriarty et al, 2018)

This activity included a stakeholder workshop held in Wellington in May 2018, where the need for this project and exploration of how it could be achieved was discussed (Moriarty et al, 2018). Work stream D included a large-scale study plan akin to the 1998-2000 FMRP, with detailed costings.

Further to receiving the report on the Survey Design and QMRA Project Plan (Moriarty et al, 2018), the Ministry for the Environment identified that a pilot study would be beneficial to confirm the size and scale of a national project. Funding was secured for this pilot (phase 1) in 2019, with sampling undertaken between February and March 2020. While the initial plan was for 80 samples, COVID19 lockdowns curtailed the sampling and analysis. Fifty-two water samples were collected and analysed from 16 rivers (initially characterised as six urban, five dairy farming and five sheep and beef farming) from around New Zealand (Leonard et al, 2020).

As well as establishing logistics, data on the prevalence and concentrations of pathogens allowed re-evaluation of the likely statistical power of a full study including the number of samples required to achieve necessary outcomes for a QMRA. Modifications were made to analytical techniques for phases 2 and 3 (Appendix B).

Funding was obtained in 2020 for the collection and analysis of 256 samples. This phase 2 sampling was undertaken over the summer – autumn period of 2021 (January to May 2021) and included monitoring of sites of specific interest to iwi and hapū.

Funding for phase 3 was secured in 2022 as part of the Jobs for Nature Programme. The sampling occurred from September 2022 to June 2024 and included one full winter and two periods of spring, summer and autumn sampling.

1.4 AIMS OF THE STUDY

The Freshwater QMRA study had four aims:

- To generate quantitative data on the concentrations of a range of pathogenic microorganisms and indicator microorganisms from a nationally representative survey of sites that were geographically spread, covered a range of climatic conditions, and had adjacent land uses or activities likely to impact freshwater quality.
- To supplement these data with information on potential contamination sources (faecal source tracking) and ancillary environmental data from sampling sites.
- To determine the relationships between pathogen and indicator presence and concentration.
- To incorporate the survey data into a QMRA providing human health risk estimates from recreational activity exposures to pathogens.

1.5 REPORT OUTLINE

Part A of this report describes the freshwater survey, and analytical results for indicators, and pathogens.

Part B provides detailed analysis of the survey data.

Part C uses the data from Part A and B in a QMRA for swimming in freshwater with different concentrations of indicators, and with different sources of faecal contamination.

Part D provides a discussion and conclusions, including a comparison of this project with the 1998-2000 project.

Appendices are used to provide supplementary information on the methodology, environmental data, analyses and data used to support the outcomes and conclusions of the project.

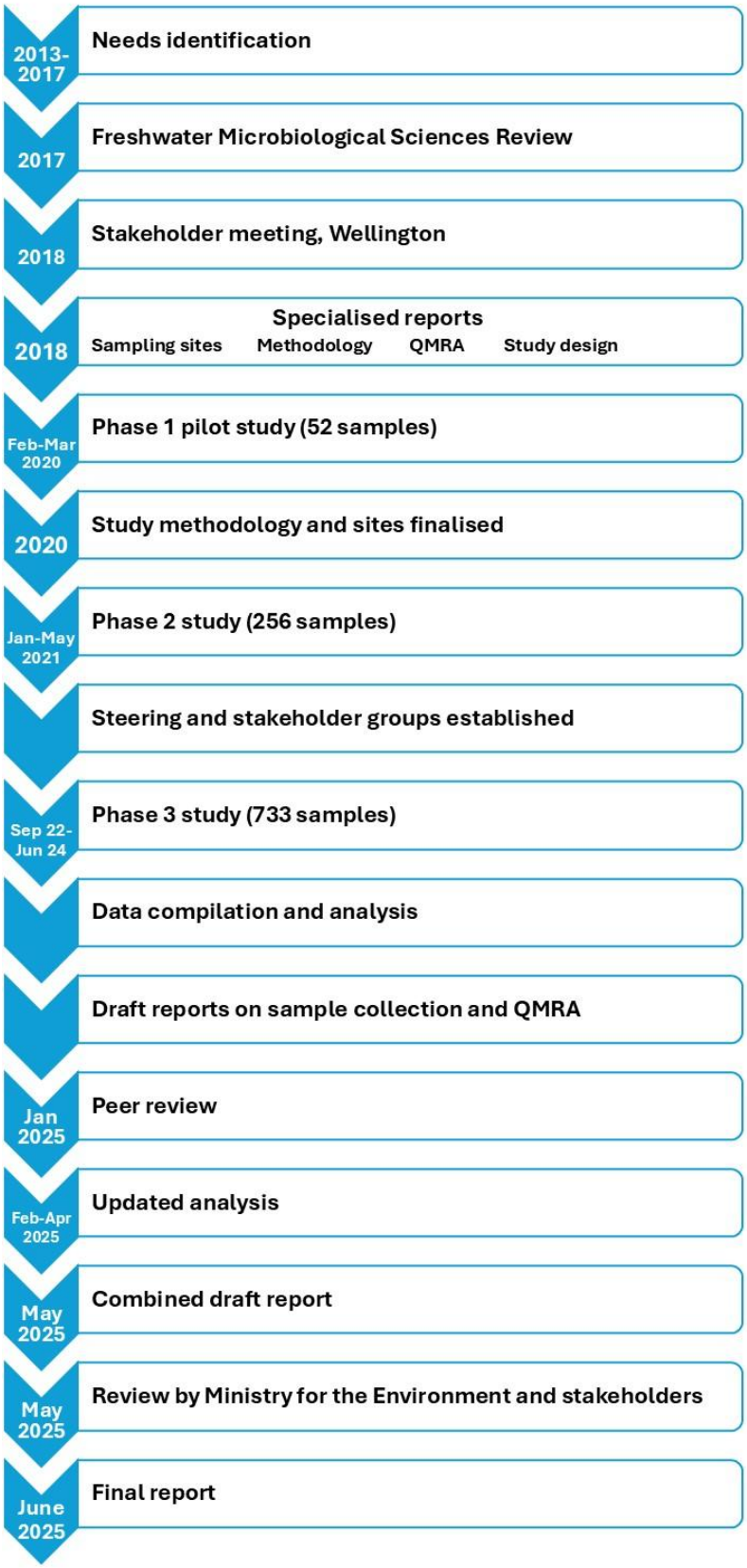


Figure 1: Study timeline

PART A

Freshwater Survey

2 FRESHWATER SURVEY INTRODUCTION

As part of the work stream B section of the Freshwater Microbiological Sciences Review (FMSR), Gilpin et al (2018) considered a variety of potentially pathogenic microorganisms and indicators which could be included in this study. Methods for detection of these microbes and recommended analytical approaches were also taken into consideration. The indicators and pathogens chosen for this study are briefly described below.

2.1 INDICATORS OF FAECAL POLLUTION

Indicator microorganisms are not usually pathogenic themselves but rather are typically found in the intestinal tract of warm-blooded animals and are therefore indicative of faecal contamination and the potential presence of pathogens.

2.1.1 Total coliforms, *E. coli* and enterococci

The most common indicators of faecal contamination used for recreational water quality monitoring are *E. coli* and enterococci, which have been used to assess the quality of recreational waters for over a century (Korajkic et al, 2018). Collectively, these bacteria are termed faecal indicator bacteria (FIB). These microorganisms are quick, cheap and easy to analyse in comparison to pathogen analysis, and there are robust, standard methods for their detection to provide consistency of results.

Recreational freshwater guidelines in New Zealand are based on levels of *E. coli*, as are the drinking water standards (New Zealand Government, 2022). *E. coli* is also the primary indicator used in other countries to monitor recreational water quality (United States Environmental Protection Agency, 2014). In contrast, marine recreational water guidelines in New Zealand are based on levels of enterococci (Ministry for the Environment, 2003). Enterococci are believed to survive better in saline water, hence their use in marine waters. There is no clear guidance on estuarine water, and several councils test for both *E. coli* and enterococci, particularly where marine beaches are impacted by rivers.

Total coliforms are a broader group of microorganisms, of which *E. coli* are one member. Assays for *E. coli* typically also generate counts of total coliforms. Compared with *E. coli* and enterococci, total coliforms have a wider range of environmental sources.

While the presence of the indicator bacteria *E. coli* and enterococci in the gut of all warm-blooded animals makes them useful indicators of faecal pollution, it doesn't provide information on the source of that faecal contamination. Different sources can pose different risks to human health and understanding these is also important for more targeted mitigations. *E. coli* and enterococci can also, under certain conditions, grow in the environment, persisting in soil or other matrices, and have non-faecal sources and reservoirs (Byappanahalli et al, 2012; Girones and Bofill-Mas, 2013).

2.1.2 Microbial source tracking markers

In this study, microbial source tracking (MST) markers were employed to better identify faecal sources. This technology was developed in the early 2000s using polymerase chain reaction (PCR) techniques to target the DNA of gut microorganisms associated with a particular host animal species. In addition, actual host DNA (e.g., mitochondrial DNA of an animal) can be targeted for specific animal identification when faecal sources are detected. More than 40 MST markers have since been developed (World Health Organization, 2021). MST markers are recognised in the WHO and Canadian recreational water quality guidelines as valuable tools to help identify sources of faecal contamination (Health Canada, 2012; World Health Organization, 2021).

While MST marker specificity is much improved compared to the non-specific *E. coli* and enterococci indicators, specificity of MST markers to a particular animal host is not absolute. They may differ in terms of:

- Sensitivity: the proportion of target host faecal specimens that test positive
- Specificity: the proportion of non-target animal faecal specimens that test positive

The frequency of detection within an animal species, and in non-target species, may differ across geographical locations (Ballesté et al, 2021; Devane et al, 2013; Green et al, 2012). This relates to different interactions between animals and humans, including co-habitation and different feed and farming practices. Cross-reaction of MST markers may occur between domestic pets and humans, across groups of farmed animals, or across animals with similar digestive physiology egg, humans and pigs.

In this study, MST markers indicative of general, human, ruminant and wild bird sources of faecal contamination were used. As human faecal contamination is potentially of most concern, and as MST markers have the potential for cross-species reaction, three human MST markers were chosen for this study. The first of these, HF183, targets bacteria in the *Bacteroides* genus. The second, BiAdo, targets the bacterium *Bifidobacterium adolescentis*. The third of these targets the bacteriophage crAssphage. All three of these markers are present in higher concentrations in human sewage than faecal indicator bacteria (FIB) (Ahmed et al, 2011; García-Aljaro et al, 2019).

HF183 is a genetic marker from *Bacteroides dorei,* which has been used extensively to characterise the human faecal input from raw sewage. It has high sensitivity, being detected in low concentrations of sewage, and has high host specificity, although it has been detected in quantifiable concentrations in an Australian study of cat and chicken faeces but not in cow, deer, goat, horse, pig or sheep (Ahmed et al, 2019). In New Zealand, cross reaction with possum, a national pest, has been reported (Devane et al, 2013). Testing of at least 200 faecal samples in New Zealand found HF183 in faeces from humans and possums, but not from cows, sheep, deer, goat, horse, dog or wild birds (ESR, unpublished data).

BiAdo is present in high concentrations in human faeces (Matsuki et al, 2004). However, it may also be present in faeces from seagulls, possums, dogs, ducks and swans, as testing of at least 200 faecal samples collected in New Zealand identified BiAdo in faeces from humans and seagulls, and low concentrations in samples from possums, dogs, ducks and swans (ESR, unpublished data).

CrAssphage is a bacteriophage, which is a virus that infects bacteria, specifically *Bacteroides* species. In human faeces, some bacteriophage are present in concentrations 10-100 times higher than gut bacteria (García-Aljaro et al, 2019). CrAssphage has high specificity for human faecal contamination and has been identified in high concentrations in New Zealand wastewater treatment plant influent and effluent (Gyawali et al, 2021). Testing of New Zealand animal and avian faeces showed that the crAssphage marker was detectable infrequently in cat and seagull faeces in quantifiable concentrations but not detectable in faeces from black swans, Canada geese, chickens, cows, dogs, ducks, horses, sheep, goats or rabbits (Gyawali et al, 2021). Possum faeces were not tested so it is uncertain whether crAssphage would be detectable in this source.

GenBac3 MST marker targets the *Bacteroidales* and is identified in concentrations in the human colon of 10^{11}/ml, around three orders of magnitude higher, than *E. coli* and enterococci (García-Aljaro et al, 2019). GenBac3 is a marker of general faecal contamination because it has low specificity being detected at high concentrations in human faeces and the faeces of animals such as possum, deer, goats, cows, sheep, dog, horse, pig, duck and black swans in New Zealand (Devane et al, 2013). GenBac3 can also be associated with non-faecal sources such as industrial waste.

BacR is a ruminant marker specific for *Bacteroidetes* bacteria found in the digestive tracts, and therefore faeces from, cattle, deer, chamois, sheep and goat, but is absent in faeces from humans, horses, pigs, cats, dogs, chickens, turkeys, swans and ducks (Reischer et al, 2006). In New Zealand, testing of more than 200 faeces samples identified BacR in faeces from cows, sheep, deer and goats; low concentrations in faeces from cats, geese, possums and some municipal sewage; and absence of detection in faeces from humans, pigs, horses, rabbits, ducks, swans, seagulls, chickens and dogs (ESR, unpublished data).

GFD is a marker for wild bird contamination and is associated with *Helicobacter* spp., (Ahmed et al, 2016; Green et al, 2012). This marker has been tested on New Zealand animal faeces and sewage and found to have moderate sensitivity for duck, swan, seagull, geese and chickens but undetectable in faeces from humans, cows, sheep, deer, horses, goats, pigs, rabbits, possums, cats and dogs (Green et al, 2012). The E2 marker, which is more specific for ducks (Devane et al, 2007), was used in Phase 1 of this project but found to be less useful than the broader GFD marker and therefore not used in the subsequent phases.

2.2 PATHOGENS

As per work stream B of the Freshwater Microbiological Sciences Review (Stage One) (Gilpin et al, 2018), the following pathogens were recommended for inclusion in this study:

- Bacteria: *Campylobacter*, *Salmonella*, Shiga toxin-producing *E. coli* (STEC)[1] and *Yersinia*
- Protozoa: Cryptosporidium and Giardia

[1]The most common strains of pathogenic *E. coli* are referred to as shiga toxin-producing *E. coli* (STEC), verocytotoxin-producing *E. coli* (VTEC) or enterohemorrhagic *E. coli* (EHEC). These three terms are interchangeable.

- Viruses: norovirus, enterovirus and adenovirus.

Selection of these pathogens was based on consideration of recommendations from the WHO (Pond, 2013), waterborne outbreaks, previous studies of pathogens in river water and disease incidence in New Zealand based on notified cases (Gilpin et al, 2018). Rates of relevant notifiable diseases in New Zealand for 2023 and 2024 are shown in Table 2, with campylobacteriosis the most frequently reported.

Notifications may include information on risk factors such as recreational freshwater contact, but these information fields are not consistently completed. In 2023, 35% of campylobacteriosis, 48% of giardiasis and approximately 60% of cryptosporidiosis, salmonellosis, yersinosis and STEC notifications had this question answered. Of these, the highest response rate for cases having recreational water contact was for giardiasis (~30%). For the other diseases, rates were around 20% apart from yersinosis (~16%). However, it should be noted that most of these cases have multiple possible risk factor exposures, so not all these cases will be due to recreational freshwater exposure.

Table 2: 2023-2024 New Zealand notifiable disease rates for relevant gastrointestinal infections

Disease	Notifications in 2023 (Armstrong et al, 2024)		Notifications in 2024 (Pattis et al, 2025)	
	Cases	Rate per 100,000 population	Cases	Rate per 100,000 population
Campylobacteriosis	6,092	116.1	5,801	108.7
Cryptosporidiosis	831	15.8	1,234	23.1
Yersiniosis	1,408	26.8	1,140	21.4
STEC infection	1,005	19.2	1,115	20.9
Giardiasis	898	17.1	844	15.8
Salmonellosis	827	15.8	844	15.8

With respect to viruses of interest for this study, norovirus, enterovirus and adenovirus infections are only notifiable under “acute gastroenteritis” and will only be notified if there is a suspected common source, or it is a person with an increased risk of spreading the disease such as health care or food service worker. In 2024, there were 37 norovirus, and two adenovirus cases reported under acute gastroenteritis. These three viruses were only analysed during the pilot study.

Due to budgetary limitations, complexity of analysis and prioritisation, *Yersinia* was not included in this study. This decision was validated by results of a recent New Zealand case-control study which found little evidence for waterborne sources of yersiniosis (Rivas et al, 2024).

Health effects (including sequelae and burden of disease) and disease incidence in New Zealand for the selected pathogens are given in Appendix A.

3 SITES AND SAMPLING DESIGN

3.1 SITE SELECTION

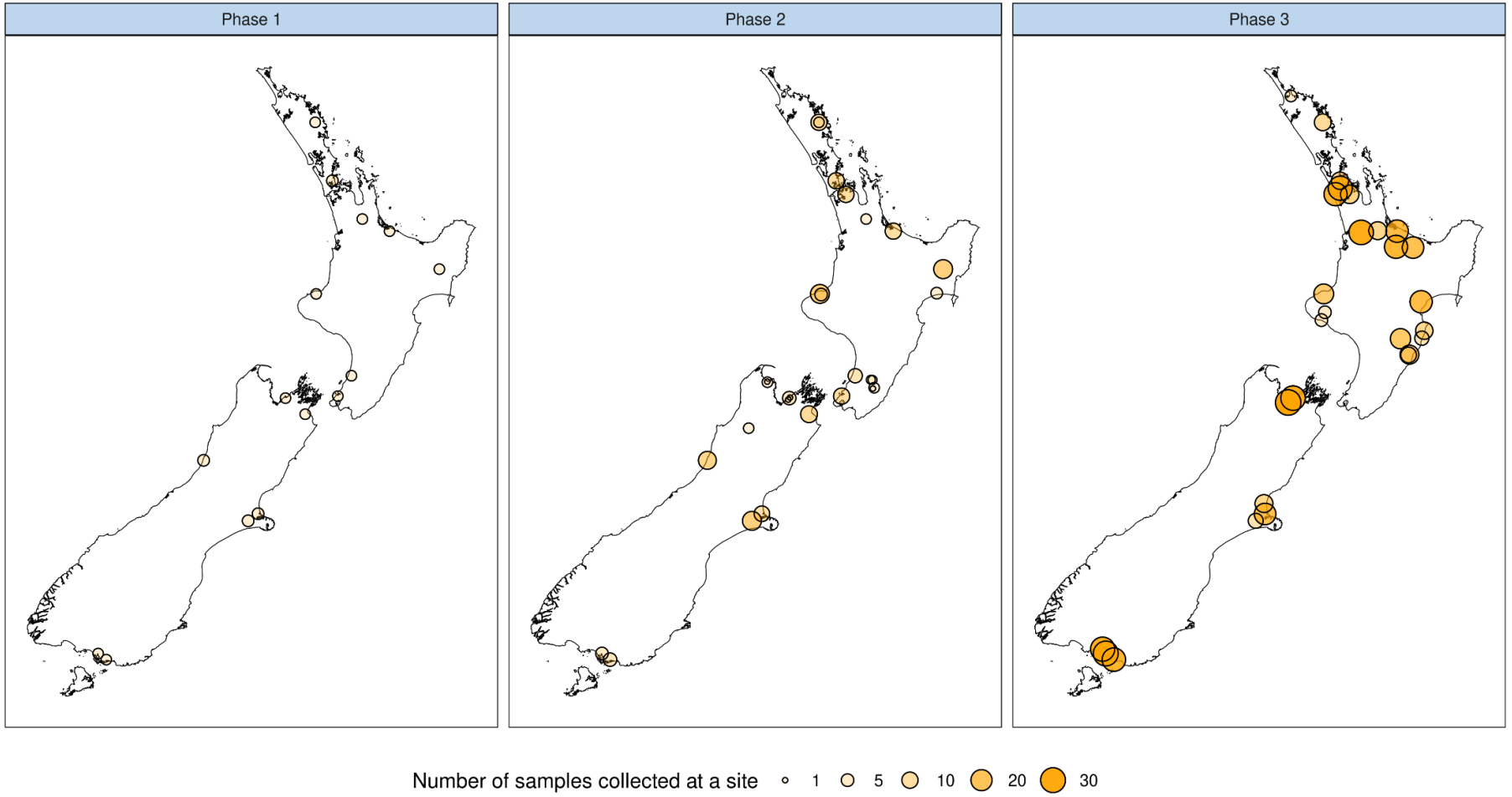
Milne et al (2018) generated a list of 30 suitable sites for the study which would provide a geographical spread, range of climatic conditions, and with adjacent land uses likely to impact water quality. To ensure sufficient data to describe relationships between pathogens and indicators, site selection emphasised sites that had a history of high *E. coli* contamination and were, therefore, likely to be impacted by faecal contamination with pathogens present.

Sampling was undertaken over four years in three phases. In each phase a different number of councils were able to participate so there were some site changes between the different phases.

- In the phase 1 pilot study, 16 sites in 13 regions were selected from the list of 30 sites identified by Milne et al (2018).
- In phase 2, 38 sites were sampled from 15 regions, which included 30 sites identified by Milne et al (2018) and another eight sites of cultural significance to iwi. These additional sites were included to support kaupapa Māori monitoring or to allow for a cultural health assessment.
- In phase 3, only nine councils were able to participate. This resulted in some previous sites no longer being sampled. However, additional sites were nominated by the participating councils resulting in a total of 46 sites being sampled.

In total, samples were taken from 71 sites, with 49 located in the North Island and 22 in the South Island (see Appendix C). Site locations are shown in Figure 2 below.

Figure 2: Site locations and number of samples collected during the three phases of this study



Note: The legend represents a scale, with examples of how the size of the circle increases with the number of samples collected at a site.

3.2 OBSERVED LAND USE CATEGORY

Councils used Land Information New Zealand (LINZ) data and internal Geographic Information Systems (GIS) to determine the different land uses in the catchment for 3 – 5 km upstream of the sampling site. These land uses were categorised according to the dominant land use with categories as follows: urban; dairy; sheep and beef; mixed dairy/sheep and beef; low impact. Where wild birds were dominant, sites were categorised as wild bird with either urban, dairy, or sheep and beef contribution (Table 3).

3.3 SAMPLES COLLECTED

In total, 1,041 water samples were collected from the 71 sites (Table 3). Samples were collected in all weathers unless there were safety concerns. Due to logistical constraints, not all sites were sampled routinely. Twenty-three sites were sampled more than 20 times (with a maximum of 38 times), 21 sites were sampled 10 – 20 times, and 27 sites were sampled less than 10 times.

Table 3: Number of samples collected across the different observed land use categories

Observed land use		Number of sites	Number of samples
Single category	Urban	12	261
	Dairy	22	236
	Sheep and Beef	16	221
	Low Impact	8	94
Multiple categories	Dairy / Sheep and Beef	2	53
	Wild bird / Urban	4	87
	Wild bird / Dairy	3	37
	Wild bird / Sheep and Beef	4	52
Total		**71**	**1,041**

4 FAECAL INDICATOR BACTERIA

Key findings

- *Samples representing a range of low to high contamination scenarios were collected including 32% of samples with E. coli concentrations less than 130 MPN/100 ml, 14% with 131-260 MPN/100 ml ,19% with 261-550 MPN/100 ml, 13% with 550-1000 MPN/100 ml and 22% with more than 1000 MPN/100 ml.*
- *Enterococci were enumerated in 85% of samples providing a second indicator bacteria for QMRA analysis.*
- *In addition to culture-based detection of E. coli, qPCR was used to detect E. coli with positive and negative predictive values of 85% compared to culture.*

Total coliforms, *E. coli*, and enterococci were enumerated in samples using culture based (Colilert, enterolert) and qPCR methods (Appendix B).

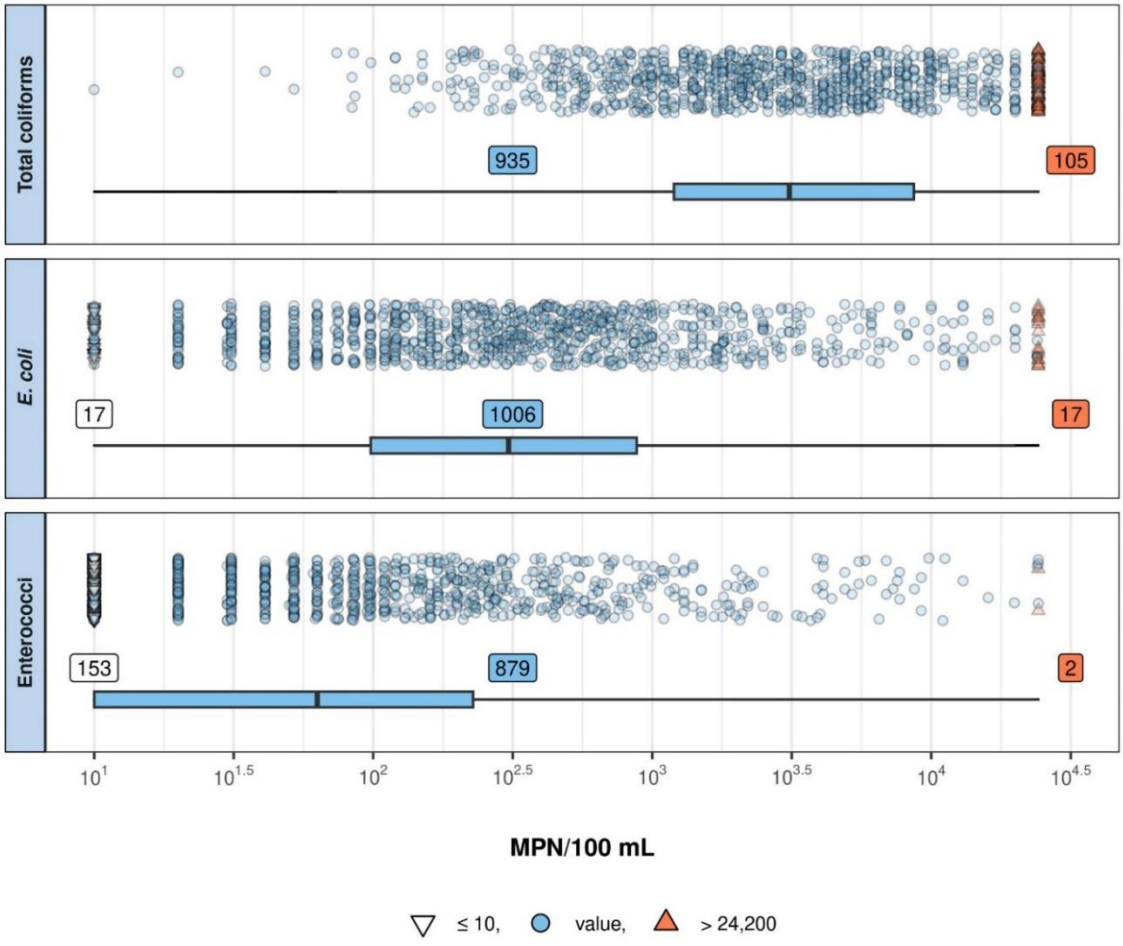
4.1 CULTURE-BASED ANALYSIS OF TOTAL COLIFORMS, *E. COLI* AND ENTEROCOCCI

Total coliforms were detected in every sample with a median concentration of total coliforms of 3,100 MPN/100 ml (Table 4 and Figure **3**). *E. coli* was present in 98% of samples with a median concentration of 310 MPN / 100 ml. Enterococci had a lower prevalence (85%) of samples with concentrations above the detection limit (10 MPN / 100 ml) with a lower median of 63 MPN / 100 ml. While median *E. coli* were 4:1 higher than enterococci, there was a wide range from 7% of samples having higher levels of enterococci than *E. coli*, through to 23% of samples with *E. coli* : enterococci ratios of 10:1 or greater. None of the faecal indicator bacteria groups were normally distributed.

Table 4: Faecal indicator bacteria concentrations by culture results

Organism		Total coliforms	*E. coli*	Enterococci
Samples analysed		1,040	1,040	1,034
Samples ≥ detection limit (10 MPN/ 100ml) (% of samples analysed)		1,040 (100%)	1,023 (98%)	881 (85%)
Samples > upper quantification limit		105	17	2
Concentration MPN/100 ml	Minimum	10	<10	<10
	25th percentile	1,200	98	10
	50th percentile (Median)	3,100	310	63
	75th percentile	8,700	880	230
	95th percentile	> 24,000	9,300	2,100
	Maximum	> 24,000	> 24,000	> 24,000

Figure 3: Concentrations of faecal indicator bacteria in collected freshwater samples



Note: The numbers in boxes indicate the number of samples with concentrations below the limit of detection (white boxes), values within the range of quantification (blue boxes) and concentrations above the limit of quantification (coral boxes). Box plots show the range of values (whisker lines), boxes represent the 25th to 75th percentile range, and line in the box the median value.

The 2003 Guidelines and 2020 NPS-FM use a number of *E. coli* concentrations to categorise freshwater samples (130, 260, 540, 550, 1,000, 1,200 *E. coli*/100 ml). There were 364 samples (35%) which exceeded the upper water quality criterion (550 MPN/100 ml) with 560 samples (54%) exceeding the water quality criterion of 260 MPN/100 ml (Table 5). The targeting of freshwater sites known to be contaminated with faecal material, has achieved the project goal of collecting samples with a good coverage of *E. coli* concentrations across the concentration range of interest.

Table 5: *E. coli* concentrations in collected freshwater samples

E. coli MPN/100 ml	Number of samples with *E. coli* concentration within range
< 130	331
131-260	149
261-540	189
261-550	196
> 540	371
> 550	364
> 1,000	229
> 1,200	212

There were 44 rivers sampled at least 10 times. Using NPS-FM (2020) Table 9 criteria to make an indicative comparison, 34 of these would be classified as E (Red), three as D (Orange), three as B (Green), and four as A (Blue) (Table 6). Thirty-nine of the rivers would be classified as Poor using Table 22 criteria of the NPS-FM, and a nationally defined bathing season, and as D grade using Microbial Assessment Categories (MAC) of the 2003 Guidelines (Table 6).

Table 6: *E. coli* based gradings of rivers sampled at least 10 times during the study for all data.

River	n	% >540	% >260	Median	95 %	Table 9 Overall Grade	Table 22 Bathing Season	MAC[1]
NI_45	10	0%	0%	10	49	A	Excellent	A
NI_30	10	0%	0%	58	180	A	Good	B
NI_14	28	4%	7%	74	410	A	Fair	C
SI_12	38	3%	8%	120	400	A	Fair	C
SI_19	17	6%	6%	52	440	B	Fair	C
SI_2	22	9%	9%	20	820	B	Poor	D
NI_27	12	8%	17%	86	880	B	Poor	D
NI_2	17	12%	24%	130	800	D	Poor	D
SI_13	13	8%	38%	200	870	D	Poor	D
SI_3	27	15%	22%	220	880	D	Poor	D
SI_4	10	10%	10%	63	1,500	E	Poor	D
SI_20	32	13%	13%	81	1,900	E	Poor	D
NI_19	14	7%	7%	87	1,600	E	Poor	D
NI_39	15	33%	33%	130	13,200	E	Poor	D
NI_10	17	18%	29%	150	2,500	E	Poor	D
SI_16	14	21%	29%	160	4,600	E	Poor	D
SI_6	21	19%	29%	190	1,400	E	Poor	D
SI_21	22	18%	18%	200	1,500	E	Poor	D
NI_4	18	22%	33%	200	1,900	E	Poor	D
NI_22	35	40%	49%	260	21,300	E	Poor	D
SI_22	28	32%	46%	270	1,700	E	Poor	D
NI_5	25	16%	28%	280	1,200	E	Poor	D
NI_38	12	42%	42%	280	12,000	E	Poor	D
NI_1	15	40%	47%	290	10,500	E	Poor	D
SI_8	35	26%	40%	300	1,300	E	Poor	D
NI_35	23	35%	43%	300	12,200	E	Poor	D
NI_16	26	35%	46%	320	18,800	E	Poor	D
NI_41	24	33%	46%	320	5,700	E	Poor	D
NI_18	21	19%	38%	320	8,700	E	Poor	D
NI_8	13	31%	77%	430	7,600	E	Poor	D
NI_36	30	43%	53%	440	4,900	E	Poor	D
NI_37	13	46%	62%	460	9,700	E	Poor	D
SI_17	27	41%	70%	500	1,500	E	Poor	D
SI_9	13	54%	69%	590	3,400	E	Poor	D
NI_40	12	67%	83%	590	13,300	E	Poor	D
NI_11	13	54%	69%	650	11,000	E	Poor	D
NI_23	35	69%	77%	680	2,500	E	Poor	D
NI_26	24	63%	83%	780	17,900	E	Poor	D
NI_34	10	80%	100%	810	1,800	E	Poor	D
NI_9	28	79%	86%	1,200	22,600	E	Poor	D
NI_7	29	83%	90%	1,400	6,700	E	Poor	D
SI_10	38	87%	92%	1,600	13,800	E	Poor	D
SI_14	17	82%	88%	1,600	15,200	E	Poor	D
NI_47	25	96%	100%	6,100	24,200	E	Poor	D

[1]Microbial Assessment Category

4.2 QPCR-BASED ANALYSIS OF *E. COLI* AND ENTEROCOCCI

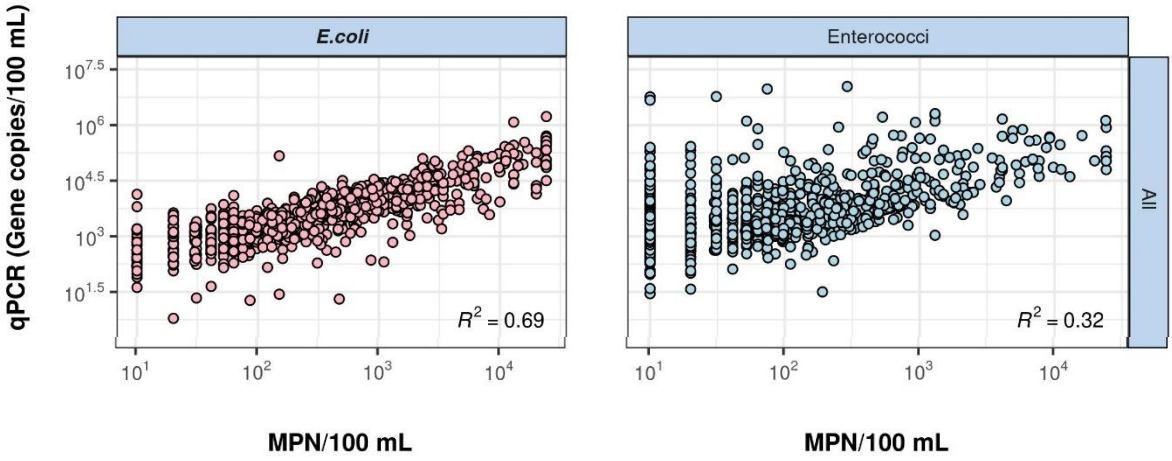
Quantitative PCR is an alternative methodology for measuring levels of *E. coli* and enterococci as gene copies (GC)/100 mL compared to culture or MPN. There were 970 samples analysed for *E. coli* and enterococci by both qPCR and culture, with these two PCR targets detected in all of the samples tested (Table 7).

Table 7: qPCR concentration estimates for *E. coli* and enterococci

Organism	Samples analysed	GC/100 ml					
		Min	25th percentile	50th percentile	75th percentile	95th percentile	Max
E. coli	970	0.4	1,580	3,870	10,700	85,600	1,690,000
Enterococci	970	28	1,230	2,790	9,860	151,000	10,960,000

A comparison of the log transformed *E. coli* and enterococci concentrations determined by culture and qPCR (Figure 4), indicates a moderate correlation for *E. coli* data (R^2=0.69) and a weak correlation for enterococci (R^2=0.32) (Schober et al., 2018).

Figure 4: *E. coli* and enterococci concentrations by culture and qPCR



These correlations were supported by comparisons of the quartile values determined by MPN culture and qPCR (Table 8, Table 9) which found 62% of the *E. coli* and 42% of the enterococci values were in the same quartile. This increases to 95% of *E. coli* and 82% of the enterococci for values in the same or adjacent quartiles. Using thresholds of above or below the median, a qPCR detection above the 50th percentile (3,870 GC/100 ml, has a positive predictive value of 85% that culturable *E. coli* will also be above the median (310 *E. coli*). Enterococci positive predictive value was 69%.

Table 8. Comparison of *E. coli* qPCR and MPN culture quartiles. If *E. coli* MPN is assumed to be the gold standard, light blue indicates potential false positives by qPCR, and dark blue potential false negatives.

***E. coli* MPN/100mL**	**qPCR gene copies / 100 ml**				
	<1,580	**1,580-3,870**	**3,871-10,700**	**>10,700**	**Total**
<10 (ND)	14 (93%)	-	-	1 (7%)	**15**
<98	152 (67%)	65 (29%)	10 (4%)	1 (0%)	**228**
98-310	60 (24%)	124 (51%)	54 (22%)	7 (3%)	**245**
310-880	14 (6%)	45 (19%)	127 (54%)	50 (21%)	**236**
>880	3 (1%)	8 (3%)	52 (21%)	183 (74%)	**246**
Total	**243**	**242**	**243**	**242**	**970**

Table 9. Comparison of Enterococci qPCR and MPN quartiles. If enterococci MPN is assumed to be the gold standard, light blue indicates potential false positives by qPCR, and dark blue potential false negatives.

***Enterococci* MPN/100mL**	**qPCR gene copies / 100 ml**				
	<1,230	**1,230-2,790**	**2,790-9,860**	**>9,860**	**Total**
<10 (ND)	83 (58%)	22 (15%)	21 (15%)	18 (13%)	**144**
10	38 (38%)	35 (35%)	15 (15%)	11 (11%)	**99**
>10-63	87 (34%)	79 (31%)	53 (21%)	35 (14%)	**254**
>63-230	30 (13%)	88 (37%)	75 (32%)	44 (19%)	**237**
>230	4 (2%)	21 (9%)	82 (34%)	137 (56%)	**244**
Total	**242**	**245**	**246**	**245**	**978**

5 SITE CLASSIFICATION BY MICROBIAL SOURCE TRACKING

Key findings

- *Microbial source tracking markers for human faecal material were identified in 35% of samples, ruminant faecal sources in 66% of samples, and wild bird faecal markers in 80% of samples. Seventy percent of samples had more than one faecal source*
- *While observed land use matched sources identified by MST analyses, there were many cases where MST identified different faecal sources to the observed sources.*
- *Use of MST markers rather than observed site classification for individual samples provides a better guide to contaminating faecal sources.*

5.1 MICROBIAL SOURCE TRACKING ANALYSES

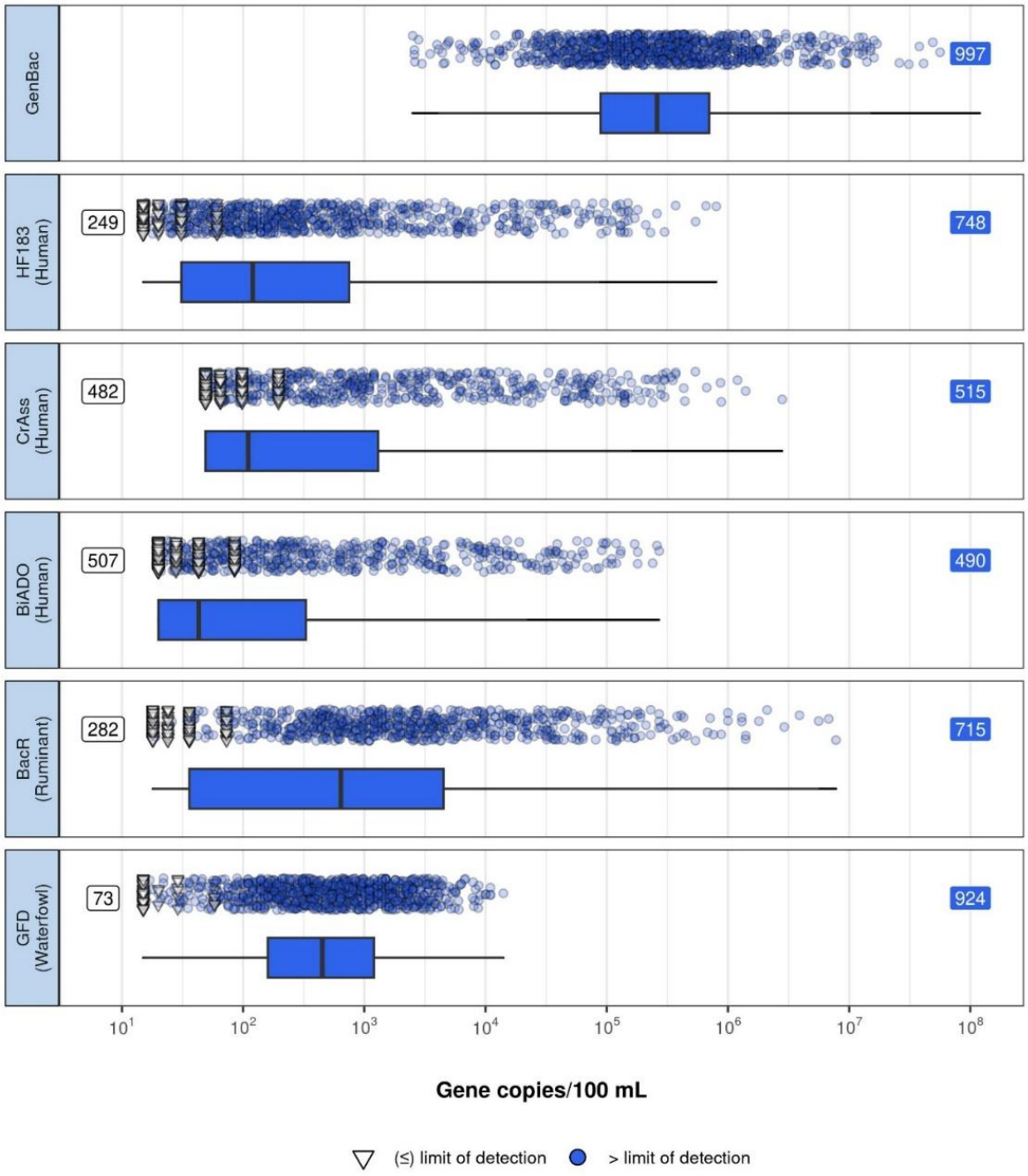
In this study, MST markers indicative of general faecal (GenBac3), human (HF183, CrAssphage and BiADO), ruminant (BacR) and wild bird (GFD) sources were used to aid in classification of the different sites. After quality control checks, data for the six MST markers detailed above was generated for 997 samples.

The general faecal marker GenBac3 was detected in every sample, and at the highest concentrations (Table 10, Figure 5). This was followed by the ruminant marker BacR, with GenBac3 and BacR concentrations ranging over 5 $\log_{10}$. Wild bird had the lowest concentration range (3 $\log_{10}$ range). The closeness of the medians to the 25th percentile for the human markers indicated there was a high number of samples with low concentrations, but the 95th percentile and maximum concentrations showed that very high concentrations of human sources occurred occasionally. As the bacteriophage, crAssphage, has a slower die-off rate than the bacterial target of HF183, and the two markers occur in sewage in similar concentrations, it would be expected that for recent, or ongoing, human faecal contamination the number of detections of HF183 and crAssphage would be similar. If the human sewage contribution was aged, then it would be expected that crAssphage concentrations would be higher than HF183 and BiADO. HF183 has also been detected in the possum population in New Zealand, likely accounting for the higher detection of HF183 compared to crAssphage and BiADO. The proportion of crAssphage and BiADO detections were similar.

Table 10: Microbial source tracking marker prevalence and gene copy (GC)/100 mL concentration in the collected water samples

Source		General	Human			Ruminant	Wild bird
MST Marker		GenBac3	HF138	CrAssphage	BiADO	BacR	GFD
Samples analysed		997	997	997	997	997	997
Samples ≥ detection limit (%)		997 (100%)	748 (75%)	515 (52%)	490 (49%)	715 (72%)	924 (93%)
Concentration GC/100 ml	Minimum	2,500	<15	<49	<20	<18	<15
	25th percentile	89,000	30	<49	<20	36	160
	50th percentile	260,000	120	110	43	640	450
	75th percentile	700,000	750	1,300	330	4,500	1,200
	95th percentile	5,300,000	43,000	89,000	24,000	130,000	3,900
	Maximum	120,000,000	800,000	2,800,000	270,000	7,800,000	14,000

Figure 5: Box and whisker plots for microbial source tracking marker concentrations in the collected water samples



Note: Values in boxes indicate the number of samples at or below the limit of detection (white boxes) or above the limit of detection (blue boxes). Whiskers extend to the minimum and maximum quantifiable concentrations. Plot boxes cover the interquartile range (25th to 75th percentiles), with the mid-line at the median (50th percentile). Detection limits can vary for each sample analysed dependant on the volume of water that was filtered.

5.2 CATEGORISATION BY MICROBIAL SOURCE TRACKING MARKERS

To aid in the presentation of the data and in describing groups of samples with mixed MST markers, categories of markers were created. For statistical analysis later in this report, the data is also modelled and presented using concentrations.

5.2.1 Single microbial source tracking types

Samples with MST detections at, or near, detection limits are unlikely to be strongly associated with significant sources of faecal contamination. These detections also had lower likelihoods of repeatability, particularly when for some sample less than 2 L of water was able to be filtered due to high particulates. To create a more robust, reproducible threshold, the minimum concentration for a group was set at twice the maximum detection limit (HF183 < 122 GC/100 ml; crAssphage < 394 GC/100 ml; BiADO < 170 GC/100 ml; BacR < 146 GC/100 ml; GFD < 116 GC/100 ml). Samples with detections below these concentrations were classified as “no human”, “no ruminant” and “no Wild birds”.

For the subgroup of samples with concentrations equal, or greater than, the above threshold values, the 25th, 75th and 90th percentiles were calculated. For the ruminant and wild bird markers, the percentile values were used to categorise the MST markers as low concentration (<25th percentile) and medium concentration (between the 25th-75th percentile concentrations). For the wild bird marker, a high concentration category corresponded to greater than the 75th percentile concentration. For the ruminant marker, the high concentration category consisted of samples with concentrations in the 75th to 90th percentile range, and the very high concentration category consisted of samples with concentrations greater than the >90th percentile concentration. The very high concentration category for the ruminant marker was included due to the wider range of concentrations observed for BacR compared to the other MST markers.

The human categories were based on the assessment of all three of the human indicative markers (crAssphage, HF183 and BiADO). All three human markers were required to be detected at greater than twice the detection limit, but with the category based primarily on crAssphage concentrations due to the longer environmental persistence of this marker and its lower levels of cross reaction with non-human hosts. For the “high human” category, HF183 or BiADO concentrations must be over the 75th percentile threshold. A category of “aged human” was created where there were elevated detections of crAssphage relative to the other human markers, suggesting partially treated or aged sources of human faecal contamination as noted earlier. These were all at low (55 samples) or medium level of the marker (26 samples). The rules used to define the different source categories are given in Table 11, with the number of samples in each category provided in Table 12.

Table 11: Concentration ranges for categorising MST makers. Percentiles are for the set of values above twice the maximum detection limit.

Category	Description
High Human	All 3 human markers concentration at greater than twice maximum detection limit **AND** crAssphage > 44,750 GC/100 ml (above the 75th percentile) **AND** {HF183 ≥ 270 GC/100 ml (above the 25th percentile) **OR** BiADO ≥ 380 GC/100 ml (above the 25th percentile)}
Medium Human	All 3 human markers concentration at greater than twice maximum detection limit **AND** 1,100 ≤ crAssphage concentration < 44,750 GC/100 ml (between 25th and 75th percentiles)
Low Human	All 3 human markers concentration at greater than twice maximum detection limit **AND** crAssphage < 1,100 GC/100 ml (below 25th percentile) **AND** HF183 and BiADO are within 1 log of concentration of crAssphage.
Aged Human	crAssphage ≥ 394 GC/100 ml (greater than twice detection limit) **AND** {HF183 <122 GC/100 ml (twice maximum detection limit) **OR** BiADO <170 GC/100 ml (twice maximum detection limit)}
No Human	concentration <394 crAssphage GC/100 ml (less than twice maximum detection limit) There are other sources of HF183 and BiADO
Very High Ruminant	concentration ≥ 89,500 BacR GC/100 ml (above 90th percentile)
High Ruminant	13,500 ≤ concentration < 89,500 BacR GC/100 ml (between 75th and 90th percentile)
Medium Ruminant	670 ≤ concentration < 13,500 BacR GC/100 ml (between 25th - 75th percentile)
Low Ruminant	146 ≤ concentration < 670 BacR GC/100ml (less than 25th percentile)
No Ruminant	concentration <146 copies BacR detected (below twice maximum detection limit)
High Wild bird	concentration ≥1600 GFD GC/100 ml (above 75th percentile)
Medium Wild bird	310 ≤ concentration < 1600 GFD GC/100 ml (between 25th – 75th percentiles)
Low Wild bird	116 ≤ concentration < 310 GFD GC/100 ml
No Wild bird	concentration < 116 copies GFD GC/100 ml (below twice maximum detection limit)

Table 12: Categorisation of water samples by microbial source tracking marker subgroups

Human markers	Number of samples (%)	Ruminant markers	Number of samples (%)	Wild bird markers	Number of samples (%)
		Very high ruminant	66 (7)		
High human	88 (9)	High ruminant	99 (10)	High wild bird	202 (20)
Medium human	147 (15)	Medium ruminant	331 (33)	Medium wild bird	403 (40)
Low human	35 (3)	Low ruminant	164 (16)	Low wild bird	196 (20)
No human	646 (65)	No ruminant	337 (34)	No wild bird	196 (20)
Aged human	81 (8)				

5.2.2 Samples with mixed microbial source tracking markers

When the MST markers in each sample are considered collectively, categories can be assigned based on the human, ruminant, and/or wild bird sources present (Table 13)

Where a sample had more than one source, if a source was present at only low levels, and those low levels were less than 1% of the human or ruminant sources present, then the low-level source was not included in creation of categories. There were 81 samples where this was a consideration.

The MST marker presence was not always consistent across samples taken from a single site reflecting the intermittent and transient nature of water contamination. This also highlights the importance of testing multiple samples to gain an accurate understanding of contamination scenarios relevant to a site.

For 301 samples (30%), a single MST source was identified. Of these, 202 (67%) were consistent with the observed land use type, while the remainder aligned more strongly to other land use types (Data not shown). More than one MST source was observed in 663 samples (66%), reflecting that in freshwater systems there will often be multiple sources of faecal contamination.

For 114 samples (11%), only low concentrations of either ruminant and/or wild bird markers were detected, with one extra sample also having low concentrations of human marker (not shown in table). These samples covered all the different observed land use categories.

For 63 samples (6%), no MST source was identified, which is either due to an absence of faecal matter or the faecal matter is from a source not detected by the MST markers used in the study. The *E. coli* concentration of these samples ranged from <10 MPN/100 ml (not detected in 6 samples) to 5,475 MPN/100 ml, suggesting for some of these samples it is possible that a faecal source not identified by the study MST markers was present. Thirty-three of the samples with no source identified were from sites classified as low impact.

As the potential human health hazard posed by faecally-contaminated waterways varies depending on the source of contamination, each freshwater sample was classified according to the MST markers detected rather than observed land use. The categories detailed above and in Table 13 were employed as they better reflected the type and extent (high or low concentration) of the faecal contamination.

Table 13: Microbial source tracking marker groupings

MST Supergroup	MST subgroup descriptions	MST Category[a]			Subtotal	Total
		Human	Ruminant	Wild bird		
Ruminant and wild bird	Ruminant and wild bird	No	Med+	Med+	196	**230**
	Low human, ruminant, wild bird	Low	Med+	Med+	8	
	Low Aged human, ruminant, wild bird	Low	Med+	Med+	26	
Ruminant	Ruminant	No	Med+	No	104	**176**
	Ruminant and low wild bird	No	Med+	Low	72	
Human and wild bird	Human and wild bird	Med+	No	Med+	93	**105**
	Human, low ruminant and wild bird	Med+	Low	Med+	12	
Wild bird	Wild bird	No	No	Med+	71	**99**
	Low human, wild bird	Low	No	Med+	20	
	Low Aged human, wild bird	Low	No	Med+	8	
Human, ruminant and wild bird	Human, ruminant and wild bird	Med+	Med+	Med+	76	**78**
	Other	Med+	Med+	No/Low	2	
Human	Human	Med+	No	No	41	**48**
	Human, low wild bird	Med+	No	Low	7	
Low ruminant and wild bird	Low ruminant and wild bird	No	Low	Med+	47	**47**
Low ruminant	Low ruminant	No	Low	No	43	**43**
Low ruminant and low wild bird	Low ruminant and low wild bird	No	Low	Low	36	**36**
Low wild bird	Low wild bird	No	No	Low	35	**35**
Aged human	Aged human	Med	No	No	7	**24**
	Aged human, ruminant, wild bird	Med	Low	Low	10	
	Aged human, wild bird	Med	No	Low	7	
Other		-	-	-		13
No source identified	No source identified	No	No	No	63	**63**

[a]Medium and high categories are combined (Med+). [b] 7 samples not included in these group combinations. [c] values in brackets indicate the number of sites the associated samples were collected from.

6 ENVIRONMENTAL DNA ANALYSIS

Key findings

- *Environmental DNA (eDNA) from 19 different mammals were detected with cattle, rat, possum, sheep, and deer the most frequently observed.*
- *eDNA from 40 different wild birds were detected with mallard duck and pūkeko the most frequently observed.*
- *These results highlight the need to consider a wide range of potential sources of faecal contamination, and a need to understand the pathogens and indicators associated with these under-evaluated non-livestock faecal sources.*

Environmental DNA (eDNA) is genetic material that is shed by organisms as they move in, through and around their environment. A number of commercial services are available in New Zealand which extract DNA from water samples and test that DNA using metabarcoding and sequencing of regions of DNA from specific species of fish, macroinvertebrates, birds, mammals, reptiles, amphibians, plants, fungi, protists, and bacteria.

In this study we evaluated the use of eDNA analysis to identify specific species of mammals and birds which may contribute to faecal contamination. Environmental DNA (eDNA) analysis was successfully completed on 289 different samples from 52 rivers with between 1 and 13 samples from each river (average 6) (Appendix A).

6.1 MAMMALS

DNA from 19 different animals was detected (Table 14). The most frequent detection was of cattle DNA detected in 51/52 rivers, and 71% of samples tested, while sheep DNA was detected in 28 rivers and 24% of samples, and deer DNA in 30 rivers and 21% samples.

Goat, horse, pig, dog and cat DNA were detected in a number of rivers, with horse and dog detected in 10% of samples.

Possum DNA was detected in 43 rivers and 43% of samples tested, with relatively high levels detected in some samples. There were 100 samples with HF183 detected without crAssphage, usually at low levels and 51 of these samples had possum DNA detected.

Rat DNA from Black and/or Norway Rat was detected in 43 (83%) of rivers at least once, and in 141 (49%) of samples.

Mouse, rabbit and hare DNA was observed in a third of rivers, but less than 10% of samples. Skink, hedgehog, stoat and ferret DNA was even less frequently observed and at low levels. While these may not be significant contributors to faecal loads, this data is useful for understanding pest management of these species.

Table 14: Animal eDNA detected

Animal	Number of rivers (%)	Number of samples (%)			
		Gene Copies/100 ml			
		Detected	20 - 100	100 - 999	>1000
Cattle	51 (98%)	204 (71%)	24 (8%)	70 (24%)	110 (38%)
Sheep	28 (54%)	68 (24%)	10 (3%)	42 (15%)	16 (6%)
Red Deer	29 (56%)	49 (17%)	12 (4%)	25 (9%)	12 (4%)
Fallow deer	10 (19%)	15 (5%)	3 (1%)	12 (4%)	0
Goat	6 (12%)	8 (2.8%)	2 (1%)	3 (1%)	3 (1%)
Horse	29 (56%)	45 (16%)	11 (4%)	28 (10%)	6 (2%)
Pig	13 (25%)	17 (6%)	6 (2.1%)	9 (3.1%)	2 (0.7%)
Dog	19 (37%)	31 (11%)	11 (4%)	15 (5%)	5 (1.7%)
Cat	16 (31%)	21 (7%)	9 (3.1%)	8 (2.8%)	4 (1.4%)
Common brushtail possum	43 (83%)	124 (43%)	37 (13%)	58 (20%)	29 (10%)
Black rat;	39 (75%)	87 (30%)	26 (9%)	49 (17%)	12 (4%)
Norway rat	37 (71%)	93 (32%)	26 (9%)	56 (19%)	11 (4%)
House mouse	18 (35%)	23 (8%)	11 (4%)	10 (3%)	2 (0.7%)
Rabbit	14 (27%)	17 (6%)	8 (2.8%)	7 (2.4%)	2 (0.7%)
Brown hare	14 (27%)	19 (7%)	10 (3%)	8 (2.8%)	1 (0.3%)
Plague skink; rainbow skink	5 (9.6%)	6 (2.1%)	1 (0.3%)	4 (1.4%)	1 (0.3%)
Hedgehog	7 (13%)	7 (2.4%)	4 (1%)	3 (1%)	0
Stoat	2 (3.8%)	2 (0.7%)	1 (0.3%)	1 (0.3%)	0
Ferret	1 (1.9%)	1 (0.3%)	1 (0.3%)	0	0
Total tested	**52**	**289**			

When compared to MST analysis, 79% of the samples with cattle or sheep DNA detected by eDNA analysis at levels above 100 copies/100 ml, also had ruminant BacR marker detected (Table 15). Among the 20 samples with ruminant MST marker detected at medium or higher level, but no cattle or sheep eDNA, 8 had horse or deer eDNA (deer are ruminants). Among the other 12 samples, seven had possum eDNA detected, and for the other 5, only one had mammalian eDNA detected (rats, dog, hedgehog). It should be noted that identification of an animal's host eDNA does not necessarily equate to faecal contamination. For example, dog food can contain horse meat and the eDNA would identify horse where no faecal contamination from horses is present. Therefore, it is important to confirm faecal contamination using MST markers that target the microbes present in a host's faeces.

Table 15: Comparison of eDNA from cattle and sheep with Ruminant BacR MST marker

MST Ruminant Marker	eDNA cattle and/or sheep marker GC/100 ml				Samples Tested
	ND	<100	≥100 to <1000	>1000	
No Ruminant	36 (47%)	25 (32%)	11 (14%)	5 (6%)	77
Low Ruminant	16 (36%)	16 (36%)	7 (16%)	6 (13%)	45
Medium Ruminant	16 (16%)	35 (34%)	39 (38%)	12 (12%)	102
High Ruminant	4 (7%)	5 (9%)	10 (18%)	38 (67%)	57
Total	**72 (26%)**	**81 (29%)**	**66 (23%)**	**61 (22%)**	**281**

6.2 WILD BIRDS

DNA from 40 different wild birds were detected (Table 16). The most common detection was of Mallard duck which was identified in 51/52 rivers and 91% of river samples, followed by Pūkeko in 42 rivers and 48% samples, and Paradise Shelduck in 29 rivers and 20% samples. Other species detected at least nine times are listed in Table 16. High levels of most wild birds were detected in at least some samples.

DNA from Blackbird, Yellowhammer, Grey warbler, Black swan, and chicken were detected in samples from 3 to 5 different rivers. DNA from blue duck, Greylag goose, Australasian shoveler, Morepork, Bellbird, Muscovy duck, Turkey, Royal spoonbill, Black-billed spoonbill, and White-faced heron were detected in only one or two samples.

Table 16: Levels of bird eDNA detected

Bird	Rivers (%)	Number of Samples eDNA copies/100 ml (%)			
		Detected	20 -100	>100	>1000
Mallard duck	51 (98%)	263 (91%)	3 (1%)	42 (15%)	218 (75%)
Pūkeko	42 (81%)	140 (48%)	7 (2%)	64 (22%)	69 (24%)
Paradise shelduck	29 (56%)	57 (20%)	11 (4%)	30 (10%)	16 (6%)
Silvereye	28 (54%)	52 (18%)	17 (6%)	32 (11%)	3 (1%)
Common starling	26 (50%)	38 (13%)	11 (4%)	26 (9%)	1 (0%)
House sparrow	24 (46%)	42 (15%)	8 (3%)	28 (10%)	6 (2%)
Song thrush	22 (42%)	27 (9%)	8 (3%)	14 (5%)	5 (2%)
Common chaffinch	21 (40%)	29 (10%)	8 (3%)	17 (6%)	4 (1%)
Pigeon	20 (38%)	48 (17%)	16 (6%)	23 (8%)	9 (3%)
Little shag	20 (38%)	35 (12%)	11 (4%)	20 (7%)	4 (1%)
Black shag	19 (37%)	31 (11%)	12 (4%)	14 (5%)	5 (2%)
New Zealand scaup; black teal	13 (25%)	24 (8%)	2 (1%)	19 (7%)	3 (1%)
Tui	13 (25%)	16 (6%)	4 (1%)	8 (3%)	4 (1%)
Brown or grey teal	10 (19%)	10 (3%)	0	9 (3%)	1 (0%)
Welcome swallow	9 (17%)	12 (4%)	5 (2%)	6 (2%)	1 (0%)
Sacred kingfisher	9 (17%)	12 (4%)	2 (1%)	10 (3%)	0
Spotted dove	9 (17%)	11 (4%)	6 (2%)	4 (1%)	1 (0%)
Goldfinch	8 (15%)	11 (4%)	3 (1%)	8 (3%)	0
Eurasian coot	7 (13%)	9 (3%)	2 (1%)	5 (2%)	2 (1%)
Canada goose	5 (10%)	9 (3%)	7 (2%)	0	2 (1%)

While some wild birds were distributed across the country, others were found in only some areas or rivers. For example, Canada geese were primarily identified in rivers in Canterbury, and Dunnock only in the South Island, and pheasant only in the North Island (Table 17).

Table 17: Regional distribution of wild bird eDNA

Wild bird	Region									
	Northland	Auckland	Waikato	Bay of Plenty	Hawkes Bay/Gisborne	Taranaki	Nelson/Tasman	Canterbury	Southland	Total
Mallard duck	20	25	28	28	30	38	19	45	28	**263**
Pūkeko	8	17	20	25	10	20	14	14	11	**140**
Paradise shelduck	3	2	2	2	11	11	5	16	5	**57**
Silvereye	2	8	5	7	3	7	5	9	6	**52**
Pigeon	1	5	4	3	3	12	1	17	2	**48**
House sparrow	1	7	8	2	4	4	1	10	5	**42**
Common starling	2	5	2	5	4	5	2	9	3	**38**
Little shag	2	1	4	7	1	8	2	8	2	**35**
Black shag	2	2	2	7	2	7	1	4	4	**31**
Common chaffinch	3	1	1	2	6	5	3	5	3	**29**
Song thrush	2	2	3	3	3	3	3	3	5	**27**
New Zealand scaup; black teal	2	1	1	1	1	1		15	1	**24**
Tui		7	1	3	1	3			1	**16**
Sacred kingfisher	2			2	2	3	3			**12**
Welcome swallow			1	1	2	2	1	4	1	**12**
Goldfinch				2	1	1	1	1	5	**11**
Spotted dove		5	1			2		2	1	**11**
Kereru; kererū		2	2		2	2	1	1	1	**11**
Brown or grey teal		1	1	1	2	1		3	1	**10**
Canada goose						1		8		**9**
Eurasian coot	1		3			1	1	1	1	**9**
Fantail	1		1	1	2		3			**8**
Pheasant	1	1	2	2		1				**7**
Dunnock							1	3	3	**7**
Magpie					2	1	2	2		**7**
Yellowhammer					3	1	1			**5**
Blackbird	1			1				1	1	**5**
Grey warbler				1	1			2		**4**
Black swan			1					2		**4**
Total	**54**	**92**	**93**	**106**	**96**	**140**	**70**	**185**	**90**	**934**

Of the 245 samples with GFD MST marker detected, only 5 didn't have at least one wild bird species detected by eDNA. There were 31 samples where GFD was not detected, but at least one wild bird species was detected. Mallard duck eDNA was detected in 27 of these, and eDNA from fantail, kererū, silvereye and chaffinch in the others. The GFD MST marker is derived from an Unclassified *Helicobacter* spp. bacteria present in the faecal material of wild bird (Green et al. 2012). Only the most common wild bird species (ducks, swans, geese, seagulls) have been tested for GFD. Not every individual bird tested has GFD (e.g. duck 76% of ducks tested positive), whereas every bird is likely to shed eDNA.

Table 18. Detection of MST GFD marker compared with eDNA detection of wild bird.

MST GFD	Total wild bird eDNA			
	ND	<100	≥100 to <1000	≥1000
ND	5 (14%)	5 (14%)	20 (56%)	6 (17%)
Low	3 (6%)	3 (6%)	29 (62%)	12 (26%)
Medium	2 (2%)	4 (4%)	46 (40%)	62 (54%)
High	(0%)	4 (5%)	24 (29%)	56 (67%)
Total	**10 (4%)**	**16 (6%)**	**119 (42%)**	**136 (48%)**

Table 19 Detection of MST GFD marker compared with detection of eDNA from Mallard Duck.

MST GFD	Mallard Duck eDNA			
	ND	<100	>100 - <1000	>1000
ND	9 (25%)	7 (19%)	15 (42%)	5 (14%)
Low	9 (19%)	7 (15%)	23 (49%)	8 (17%)
Medium	3 (3%)	11 (10%)	55 (48%)	45 (39%)
High	1 (1%)	3 (4%)	30 (36%)	50 (60%)
Total	**22 (8%)**	**28 (10%)**	**123 (44%)**	**108 (38%)**

There have not been any comprehensive studies of levels of pathogens and indicators in many of these species – particularly rats and most of the wild birds. Studies of other animals (cattle, sheep, deer, dogs, cats, possum) and wild bird (mallard duck) may need to be updated, and in some cases additional pathogens tested for.

These results highlight the need to consider a wide range of potential sources of faecal contamination.

7 PATHOGENS

Key findings

- *Campylobacter were frequently isolated (68% of samples), with 4% of samples having C. jejuni ≥ 100/100 ml, 23% ≥ 10/100 ml, and 42% with ≥ 1/100 ml.*
- *C. coli was detected in 13% of samples, but only 3% of samples had ≥ 1/100 ml.*
- *At least 44% of the Campylobacter genotypes have a high likelihood of causing disease in humans, while the pathogenicity of the rest is unknown as they are novel genotypes or only previously found in wild birds.*
- *Cryptosporidium and Giardia were detected in 50% and 69% of samples respectively, with 1% of samples having ≥ 10 Cryptosporidium/10 litre and 3% of samples having ≥ 10 Giardia/10 litre.*
- *Salmonella were detected in up to 19% of samples, and potential STEC in 20 – 30% of samples. Levels of each organism were almost always <1/100 ml, andh only 1% of samples yielding a sequenced STEC.*
- *Pathogenic viruses were detected at very low levels in the pilot study, so an indicator virus (CrAssphage) was evaluated as a surrogate to assess viral risk.*

This section provides a summary of the prevalence and concentrations of selected pathogens in the water samples. For *Campylobacter*, *Salmonella*, STEC, *Giardia* and *Cryptosporidium,* samples were analysed during each phase of the project by culturing, microscopy and/or qPCR (Appendix B). For norovirus, enterovirus and adenovirus, samples were assessed by qPCR during the pilot study only.

The sites used for water sampling were chosen to ensure a focus on sites with higher likelihoods of faecal contamination, and therefore the prevalence and concentrations of pathogens may be higher than might be observed in a typical or average freshwater site.

7.1 INTRODUCTION TO MULTILOCUS SEQUENCE TYPING

Isolates of *Campylobacter*, *Salmonella* and STEC were subjected to whole genome sequencing. The assembled genomes (~1.6, 3 and 5 million bases pairs each, respectively) were then analysed to facilitate more accurate identification and to allow for comparisons of the relatedness between isolates.

Multilocus sequence typing (MLST) is a technique used to characterise different isolates of bacterial species based on differences in the sequences of specific genes, or regions of genes. This technique was first developed by sequencing 450 – 500 bp of each of seven house-keeping genes that are present in every isolate of a species. Each different version of

the house-keeping gene (whether differing by one or many base pairs) is assigned a distinct allele number. Isolates are then assigned an allelic profile or sequence type (ST) based on the series of seven allele numbers (Maiden et al, 1998), with online international databases allowing for consistent naming of alleles and STs (Jolley et al, 2018). The key features of MLST are that:

- Isolates with the same ST have the same sequence at each of the seven loci sequenced.
- Isolates with a different ST differ in at least one of the loci sequenced.

Bacteria have thousands of different genes which can be sequenced concurrently using whole genome sequencing (WGS). This allows MLST analysis to be undertaken using more than seven genes, with schemes using hundreds to thousands of genes available. MLST schemes which utilise WGS data (e.g., core genome MLST (cgMLST) and whole genome MLST (wgMLST)) provide much higher resolution due to the increased proportion of the bacterial genome being sequenced.

In this report MLST results are presented in figures as minimum spanning trees. Each circle represents a different combination of alleles. If a circle is subdivided it means that there are two or more isolates which have the same sequence at every allele they have in common. The numbers on the branches between circles indicate the number of allele differences. Using the wgMLST schemes, isolates with 10 or fewer differences may be considered closely related, with a likely common source. As the number of allele differences increases, the less likely it is that two isolates have a common source.

7.2 CAMPYLOBACTER

Two 2018-2019 campylobacteriosis case-control studies conducted in New Zealand (Lake et al, 2021) and Australia (Cribb et al, 2022) typed the *Campylobacter* species associated with cases. In the New Zealand study, 95% of the 666 included cases were found to be caused by *Campylobacter jejuni* (Lake et al, 2021) and in the Australian study *C. jejuni* was responsible for 74% of the 571 study cases (Cribb et al, 2022). In both studies, the majority of remaining cases were caused by *C. coli.* Under normal circumstances, *Campylobacter* cases are not routinely typed in New Zealand.

Campylobacter jejuni and C. coli are two members of the broader group of thermophilic *Campylobacter* which contains many other species including *C. lari* (Cookson et al, 2024). The health risk associated with these other species *Campylobacter* species is largely unknown (Cookson et al, 2024).

During this study, *C. jejuni*, *C. coli* and total thermophilic *Campylobacter* were enumerated using MPN enrichment and direct PCR of DNA extracted from the water samples.

7.2.1 Isolation of Campylobacter by culture

The presence of *Campylobacter* in 1,028 water samples was assessed by culturing (Appendix B). The prevalence and concentrations of *Campylobacter* in these samples are shown in Table 20. Thirteen water samples were excluded from this assessment as they did not meet the requirements of transport time and/or temperature.

Of the 1,028 samples analysed, *C. jejuni* was detected in 64% (656), *C. coli* was detected in 13% (136), and in 12% of samples (124 total) both species were detected. In a further 3% of samples (28 total) only thermotolerant *Campylobacter* of another species was detected.

In this study, the sampling plan was designed to achieve 10% of samples with *Campylobacter* concentrations above 10 MPN/100 ml. In total, there were 240 samples (23%) with *C. jejuni* concentrations greater than 10 MPN/100 ml, with 46 of these greater than 100 MPN/100 ml, and three greater than 1,000 MPN/100 ml. Only three of the 137 samples in which *C. coli* was detected had concentrations greater than 10 MPN/100 ml (all 23 MPN/100 ml), with 80% of samples having concentrations less than 1 MPN/100 ml.

Table 20: *Campylobacter* prevalence and concentrations for all collected freshwater samples

Organism		Total thermophilic *Campylobacter*	*C. jejuni*	*C. coli*
Samples analysed		1,028	1,028	1,028
Samples ≥ detection limit (% of samples analysed)		695 (68%)	656 (64%)	136 (13%)
Samples > upper quantification limit		1	1	0
Concentration MPN/100 mL	**Minimum**	< 0.1	< 0.1	< 0.1
	25th percentile	< 0.1	< 0.1	< 0.1
	50th percentile (Median)	0.34	0.34	< 0.1
	75th percentile	14	9.2	< 0.1
	95th percentile	115	93	0.34
	Maximum	> 1,100	> 1100	23

Campylobacter species were recovered from 94% (67/71) of sampling locations, with *C. jejuni* isolated at 65 of these locations, and *C. coli* at 39 sites (Table 21). At 29 of the sites, *C. jejuni* were isolated from at least 80% of the samples collected, while there were only 9 sites where *C. coli* was isolated from 40% or more of the samples tested.

Table 21: Frequency of isolation of *C. jejuni* and *C. coli* at different sampling locations

Frequency of Detection	*C. jejuni*			*C. coli*		
	Number of Sites	Samples Tested	Samples Detected	Number of Sites	Samples Tested	Samples Detected
100%	14	95	95	2	3	3
80-96%	15	266	242	1	27	24
61-79%	14	258	179	1	3	2
40-57%	11	203	203	5	100	48
20-36%	5	88	26	8	73	18
3-19%	6	91	13	22	522	41
0%	6	27	0	32	300	0

7.2.2 *Campylobacter* qPCR analysis

Campylobacter qPCR analysis was undertaken on 945 samples. In 442 of these samples (47%) *C. jejuni* was detected at quantifiable concentrations (30 – 19,000 genome copies/100 mL, median 148 GC/100 mL). In a further 15% of samples *C. jejuni* was detected below the level of quantification. In 115 of the 945 samples (12%) *C. coli* was detected at quantifiable concentrations (36 – 1,200 GC/100 mL, median 113 GC/100 mL), and a further 8% contained *C. coli* at levels below the limit of quantification.

Of those samples where *C. coli* was detected by qPCR, 91% also had *C. jejuni* at detectable concentrations.

A total of 943 samples were analysed using MPN culture and qPCR. All of the samples with *C. jejuni* present with a MPN of 115/100 mL or greater had quantifiable concentrations of *C. jejuni* by qPCR (Table 22). If MPN-based culture analysis is taken as the gold standard, then qPCR provides a relatively low level of false negatives (9%) when the MPN is greater than 10, and 19% false negative rates when MPN is greater than 1 MPN/100 ml. Overall qPCR has a positive predictive value of 84% when greater than 1 MPN/100 ml. In samples where *C. jejuni* was not isolated by culturing, 37% had *C. jejuni* detected by qPCR, although only nine of these samples had more than 200 GC/100 ml.

Table 22: Comparison of MPN-based culture analysis with qPCR for detection of *C. jejuni*

Culture-based analysis		Quantitative PCR			
MPN/100 mL	Count	Not detected	Detected below limit of quantitation	Quantified	GC/100 mL range (median)
Not detected	339	214 (63%)	46 (14%)	79 (23%)	30 – 980 (58)
0.1 - 0.7	206	79 (39%)	44 (21%)	83 (40%)	30 – 750 (100)
3 - 10	185	46 (25%)	37 (20%)	102 (55%)	30 – 1100 (130)
14 - 93	174	17 (10%)	19 (11%)	138 (79%)	36 – 1300 (250)
115 - >1,100	39	0	0	39 (100%)	110 – 19000 (580)

Using MPN-based culture analysis *C. coli* was detected less frequently and at lower concentrations than *C. jejuni*, and qPCR analysis supported these low concentrations with very low levels of C. coli detected (Table 23). Of the samples where *C. coli* was isolated by culture, *C. coli* was detected by qPCR in 39% of these samples.

Table 23: Comparison of MPN-based culture analysis with qPCR for detection of *C. coli*

Culture-based analysis	Quantitative PCR				
MPN/100 mL	Count	Not detected	Detected below limit of quantitation	Quantified	GC/100 mL range (median)
Not detected	833	692 (83%)	59 (7%)	82 (10%)	40 – 530 (90)
0.1 - 0.4	94	59 (63%)	12 (13%)	23 (24%)	40 – 480 (200)
3 - 23	24	13 (54%)	1 (4%)	10 (42%)	110 – 1200 (180)

7.2.3 *Campylobacter* whole genome sequencing

Campylobacter isolated from 540 water samples from 60 different sites were whole genome sequenced. For 95 of the samples, more than one different isolate was sequenced. The majority of *Campylobacter* isolates sequenced were *C. jejuni* (515), *C. coli* (62), and *C. lari* (5).

In addition, 63 isolates sequenced were other *Campylobacter* species including four novel subspecies of *Campylobacter* (Table 24). These species were isolated from 27 different sites, 26 of which also had *C. jejuni* and/or *C. coli*. There was one site where the only species isolated was *Campylobacter porphyrionis* subsp. avium subs. nov. These other *Campylobacter* species have not been confirmed as pathogens of humans. As such, the QMRA is based on the levels of *C. jejuni* and *C. coli*, not total thermophilic *Campylobacter*.

Table 24: Other species of *Campylobacter* isolated from water samples

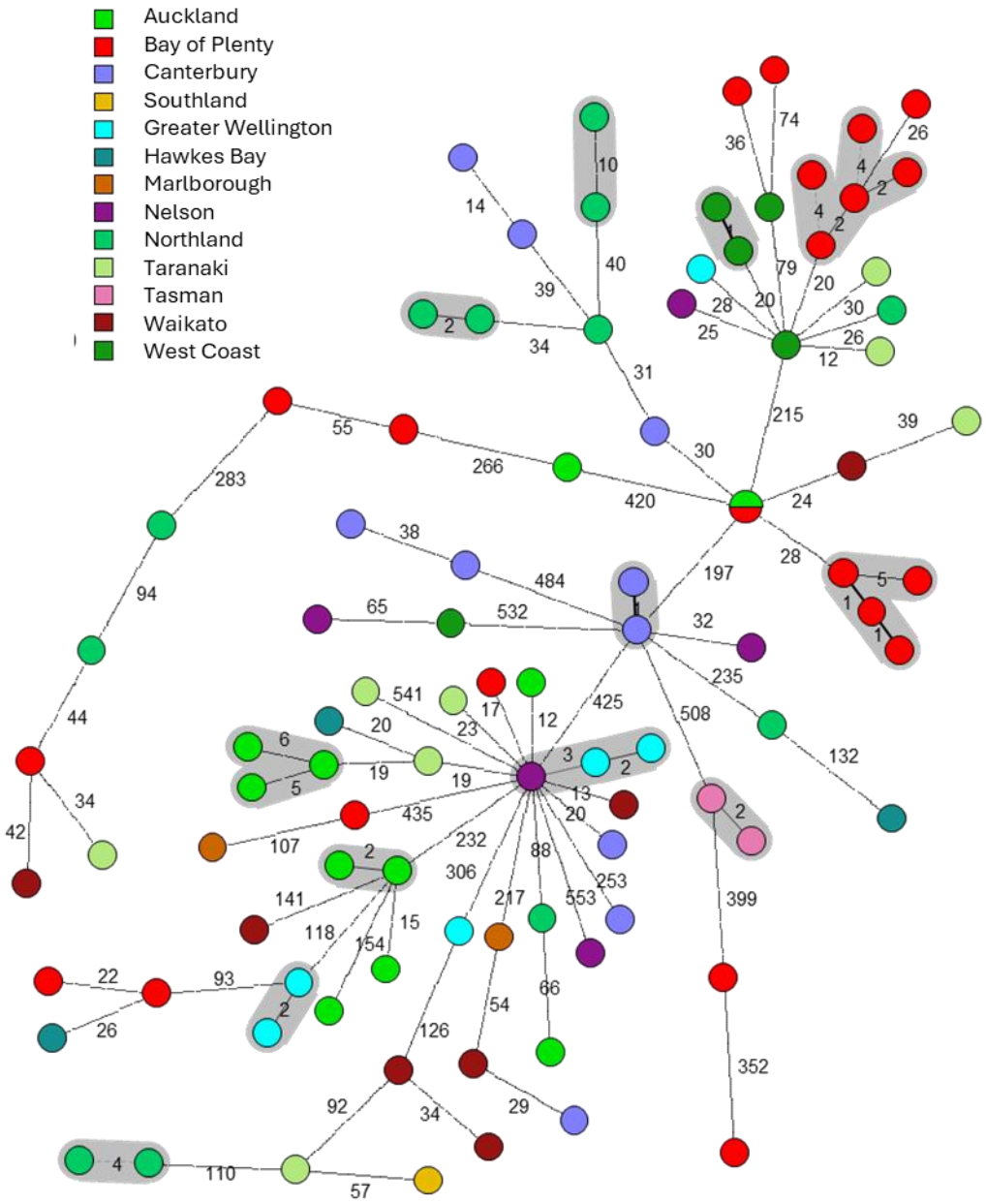
Species	Number of Isolates Sequenced	Number of Sites
***Campylobacter gallirallus* sp. nov**	2	2
Campylobacter lari	5	3
***Campylobacter porphyrionis* subsp. avium subs. nov**	6	6
***Campylobacter porphyrionis* subsp. hochstetteri subs. nov**	4	3
***Campylobacter porphyrionis* subsp. porphyrionis subs. nov**	22	14
***Campylobacter porphyrionus* subsp. notornithis subs. nov.**	14	12
***Campylobacter rallidarum* sp. nov.**	3	2
***Campylobacter rivi* sp. Nov**	4	3
Other *Campylobacter* sp.	8	4

7.2.4 *Campylobacter jejuni* whole genome sequencing types

Among the 515 *C. jejuni* isolates that were sequenced, there were 130 different seven gene sequence types (STs). Sixty-four of the 130 STs were seen only once, 23 were seen twice, and 15 were seen three times. Among these STs were 37 new STs which had not been seen before (corresponding to 54 isolates).

The most frequently observed ST was ST2381, which was observed 96 times at 31 different sites. There was significant diversity within ST2381 isolates (Figure 6), with 31 isolates (32%) in 13 clusters that could be considered to have a common source (less than 10 wgMLST differences). Within each cluster, isolates were all from the same or neighbouring region. This ST has only been previously found in wild bird.

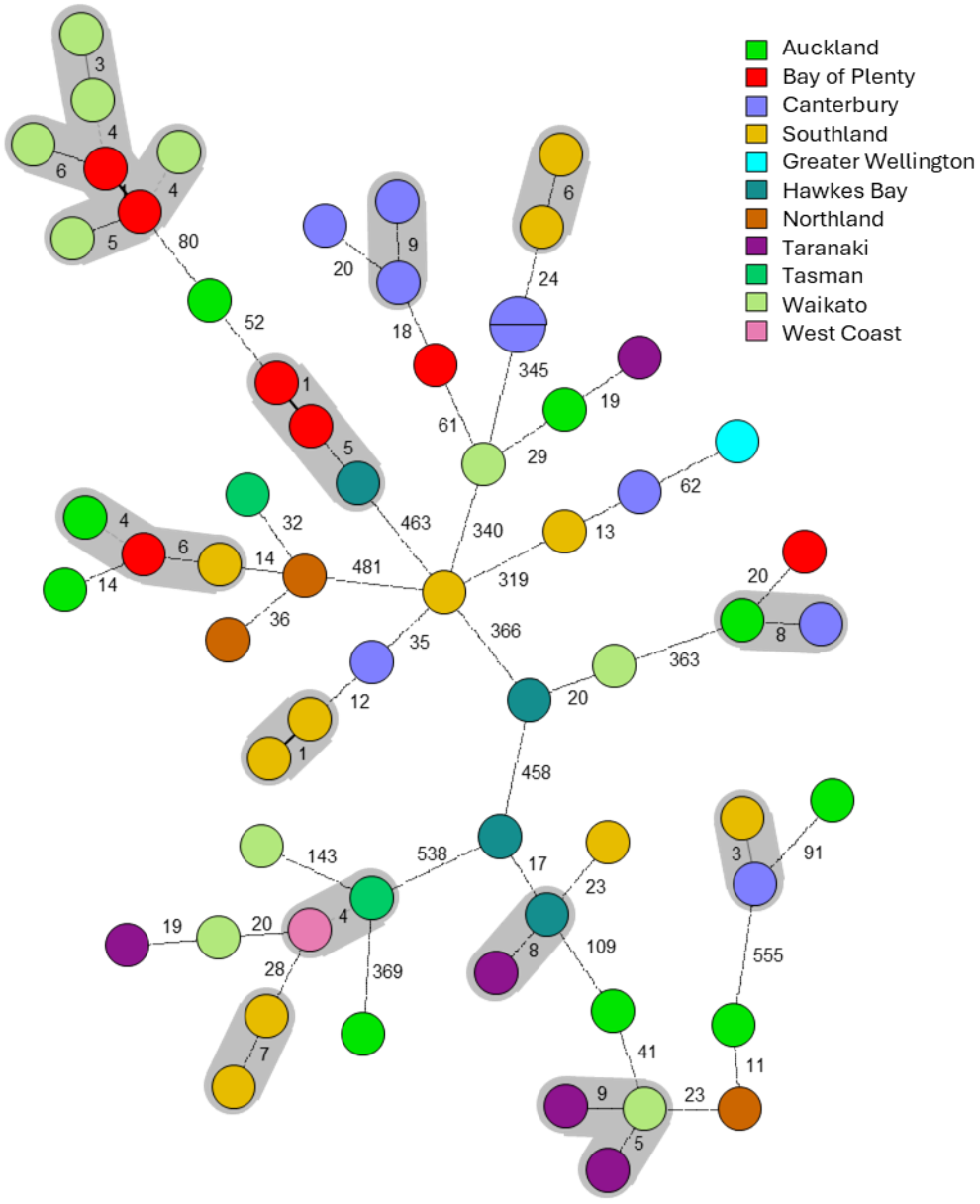
Figure 6: Whole genome MLST minimum spanning tree of *C. jejuni* ST2381 isolates



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences.

The second most frequent ST, ST45, was identified in 62 isolates from 30 sites. While again there was significant diversity, 30 of the isolates (48%) were in potential common source clusters. Isolates within these potential clusters were found in an increased number of regions compared with ST2381 (Figure 7). This ST has been frequently observed in humans, and isolated from ruminant, poultry and wild bird sources.

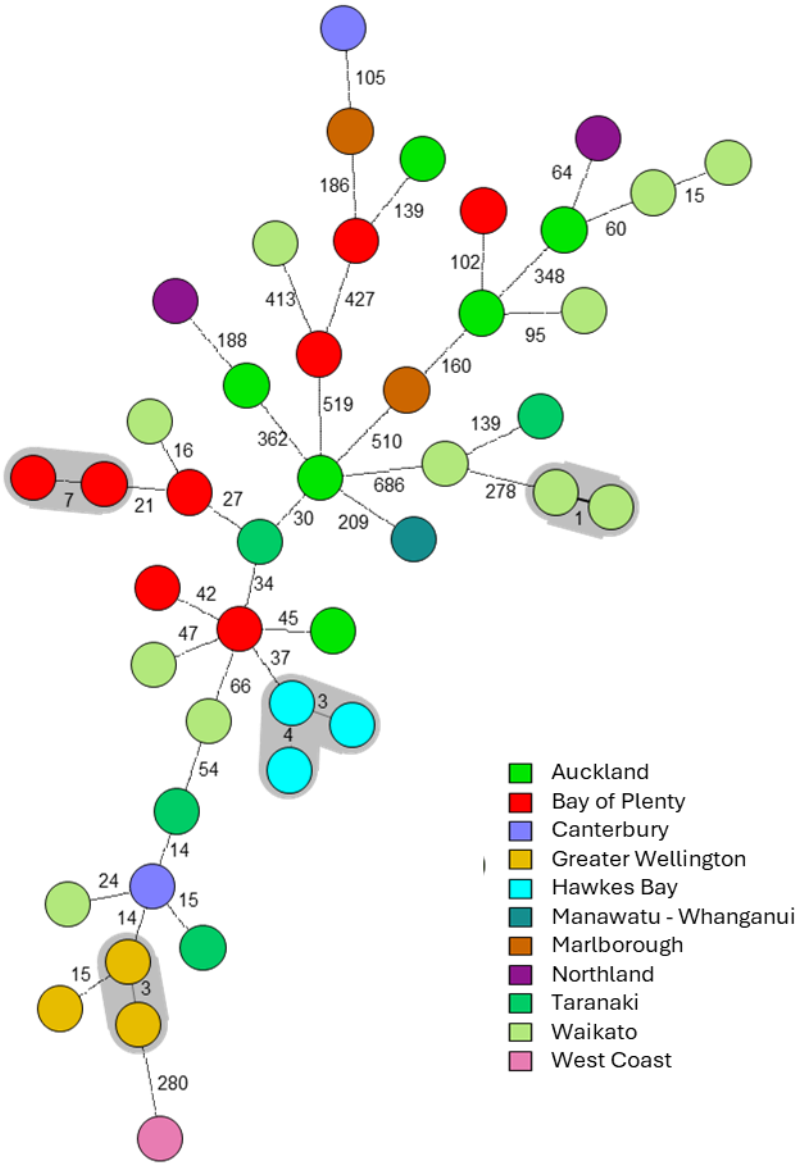
Figure 7: Whole genome MLST minimum spanning tree of *C. jejuni* ST45 isolates



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences.

The third most frequent ST, ST3655, was identified in 44 isolates from 23 sites. This ST, which has only previously been isolated from wild bird, displays significant diversity with few potential common source clusters (9 isolates (20%) in 4 clusters), which were each only from the same region (Figure 8).

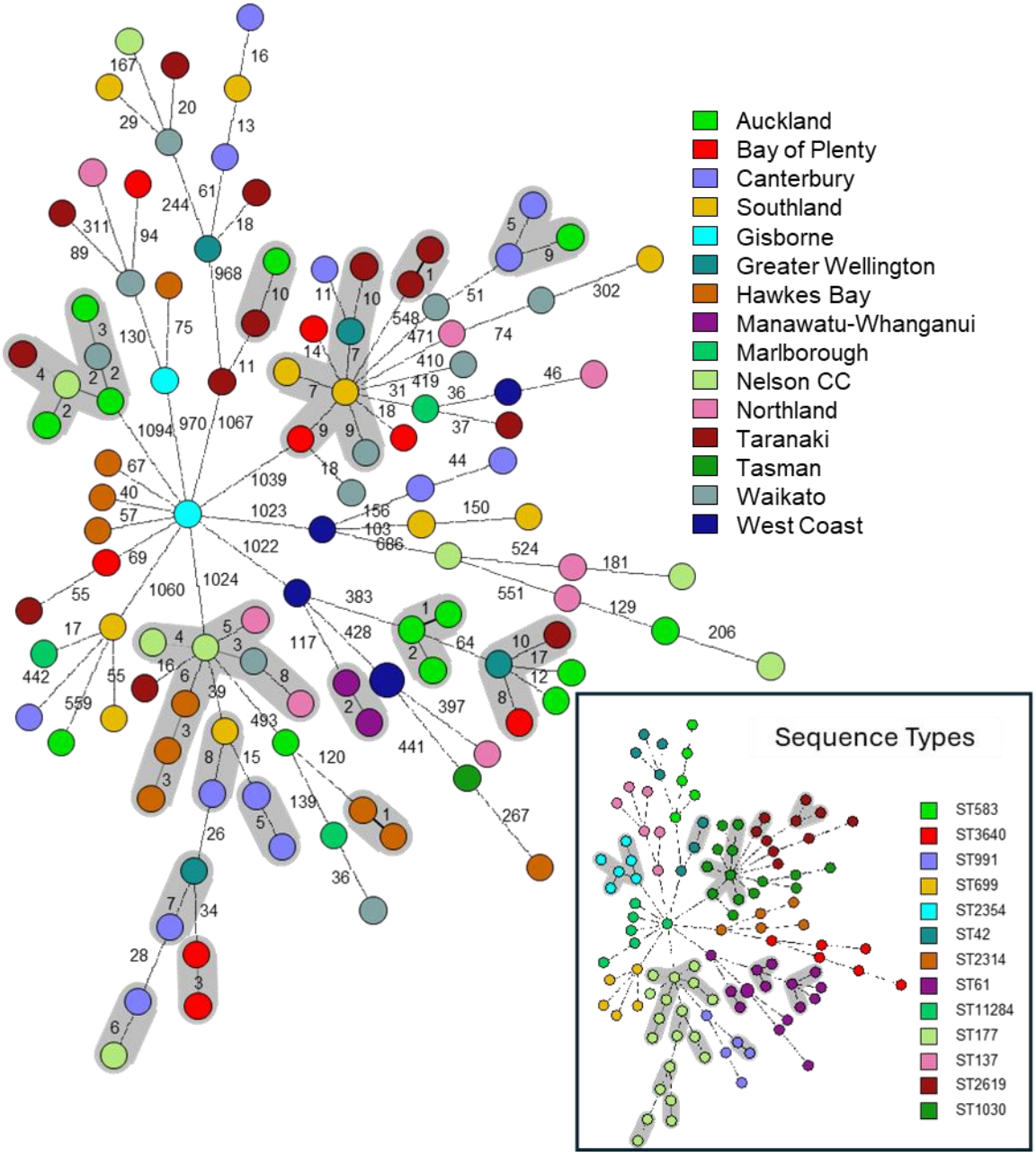
Figure 8: Whole genome MLST minimum spanning tree of *C. jejuni* ST3655 isolates



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences.

There were another 110 isolates which could be distinguished into 13 STs with 5 – 19 isolates in each ST (Figure 9). One of these STs (ST11284) was novel, two STs were only previously isolated from wild bird (ST3640 and ST2619). ST2354 has only previously been isolated from poultry and wild bird while the others have all been previously isolated from human, poultry, ruminant and wild bird. Clusters of these isolates included a large proportion from different regions.

Figure 9: Whole genome MLST minimum spanning tree of *C. jejuni* isolates with STs observed 5 – 19 times



Numbers on branches are the number of wgMLST differences. Each colour in the main diagram corresponds to a different region of New Zealand, while the box highlights 7 gene MLSTs. Grey shading indicates 10 or fewer wgMLST differences.

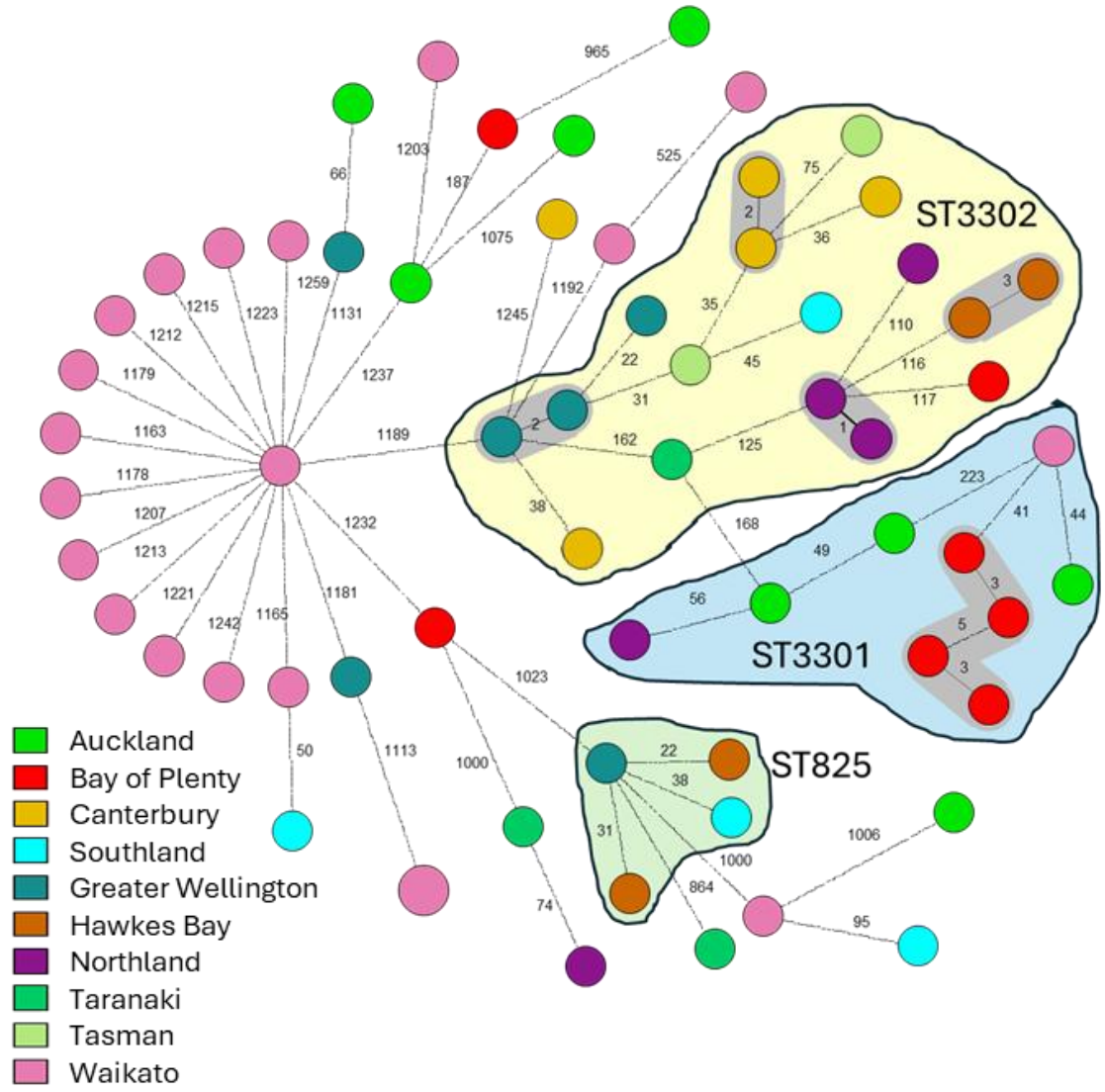
Overall, 46 of the *C. jejuni* STs have previously been observed in human clinical cases (209 isolates). Among these, 39 ST (200 isolates) were also found in poultry and/or ruminants, with 27 of these (102) also found in wild bird. Another 8 ST (19 isolates) have been found in poultry or ruminants. Together this suggests that at least 44% of the *C. jejuni* (54 ST, 228 isolates), have strong evidence for being likely to cause disease in humans.

Within the other 76 ST,14 ST (183 isolates including the 96 ST2381) have previously been isolated from wild bird, with the remaining ST novel to this study. Whether these isolates have the same human pathogenic potential as the other isolates is unknown.

7.2.5 *Campylobacter coli* whole genome sequencing types

Overall, 62 *C. coli* isolates were sequenced, corresponding to 32 different STs. The most frequently observed was ST3302, which was found 17 times at 13 different sites. ST3301 was observed 8 times, and ST825 four times. While three of the other STs were observed twice each. All observed STs were genetically very different (Figure 10). The isolates with the fewest genetic differences were all from the same regions. There were 17 new *C. coli* STs which had not been seen before (from 19 isolates). Six of the observed STs have previously been seen in human clinical cases (11 isolates) whilst three STs have only previously been reported in wild bird.

Figure 10: Whole genome MLST minimum spanning tree of *C. coli* isolates



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences. The ST3302, ST3301 and ST825 isolates are highlighted in yellow, blue and green respectively.

7.3 SALMONELLA

There are more than 2,500 different *Salmonella* serotypes (serovars), with more than 1,500 of these belonging to the *S. enterica* subspecies *enterica* group (Kingsbury et al, 2024). In New Zealand in 2024, more than 70 different serotypes were isolated from a total of 832 human infections.[2] The mostly frequently observed serotypes were Typhimurium with 261 cases (31% of total) and Enteritidis with 98 cases (12% of total). Typing information is available for these isolates, revealing that in 2024, ST19 was the dominant type for Typhimurium (125 cases) and ST11 was the dominant type for Enteritidis (68 cases).

During this study, *Salmonella* was enumerated using culture methods and qPCR, as detailed in Appendix B.

7.3.1 Isolation of Salmonella by culture

Salmonella was isolated from 164 (16%) of the 1,033 analysed freshwater samples using culture methods. MPN enumerations were only undertaken in phase 1 of this study, with nine of the 49 analysed samples positive in that phase. One sample had a concentration of 0.25 MPN/100 ml, and the other eight samples 0.21 MPN/100 ml. In phase 2 and 3, only presence/absence testing in 1 L of water was conducted.

Salmonella was detected in 63% of sites (45 total), with two sites having *Salmonella* present in more than half of their samples, as shown in Table 25.

Table 25: Frequency of isolation of *Salmonella* from water samples

Percentage of samples at a site with *Salmonella* detected	Number of sites	Samples tested	Samples with *Salmonella* present
55-67%	2	35	20
31-40%	8	120	41
20-29%	14	252	62
10-18%	12	222	30
3-9%	9	213	11
0%	26	191	0

7.3.2 *Salmonella* qPCR analysis

Quantitative PCR analysis for *Salmonella* was performed on 961 samples. In 14 of these samples (1.5%) *Salmonella* was detected at quantifiable levels (60 – 410 GC/100 ml). In another 21 samples (2.2%) *Salmonella* was detected, but below the limit of quantification (Table 26).

A total of 954 samples were analysed by both MPN-based culture analysis and qPCR. Of the 154 samples where *Salmonella* were isolated by culture analysis, only 7 (4.5%) had *Salmonella* detected by qPCR. Of the 800 samples where *Salmonella* were not isolated by

[2] https://www.esr.cri.nz/media/idcj5mvs/annual-human-salmonella-2024.pdf

culture, 11 (1.4%) had *Salmonella* quantified by qPCR (maximum of 283 GC/100 ml), while another 17 (2.1%) had *Salmonella* detected by qPCR but below the level of quantification.

Table 26: Comparison of MPN-based culture analysis and qPCR for detection of *Salmonella*

qPCR	Culture (detection in 1L)		
	Not detected	Present	Total
Not detected	772	147	919
Detected below limit of quantitation	17	4	21
60 – 91 GC/100 mL	6	1	7
150 – 410 GC/100 mL	5	2	7
Total	**800**	**154**	**954**

The MPN and qPCR results both suggest very low concentrations of *Salmonella* being present in these water samples, with perhaps no more than 4 to 7 of the samples with a concentration of *Salmonella* greater than (<1 /100 ml) (based on qPCR results).

7.3.3 Whole genome sequencing of *Salmonella*

Salmonella from 164 samples collected from 45 sites were subjected to whole genome sequencing. For 17 of these samples between one and four isolates were sequenced. Only one of these samples, had two different *Salmonella* sequenced (Seftenberg and Stanley), with the rest all identical within each sample. These duplicates were removed from subsequent analyses.

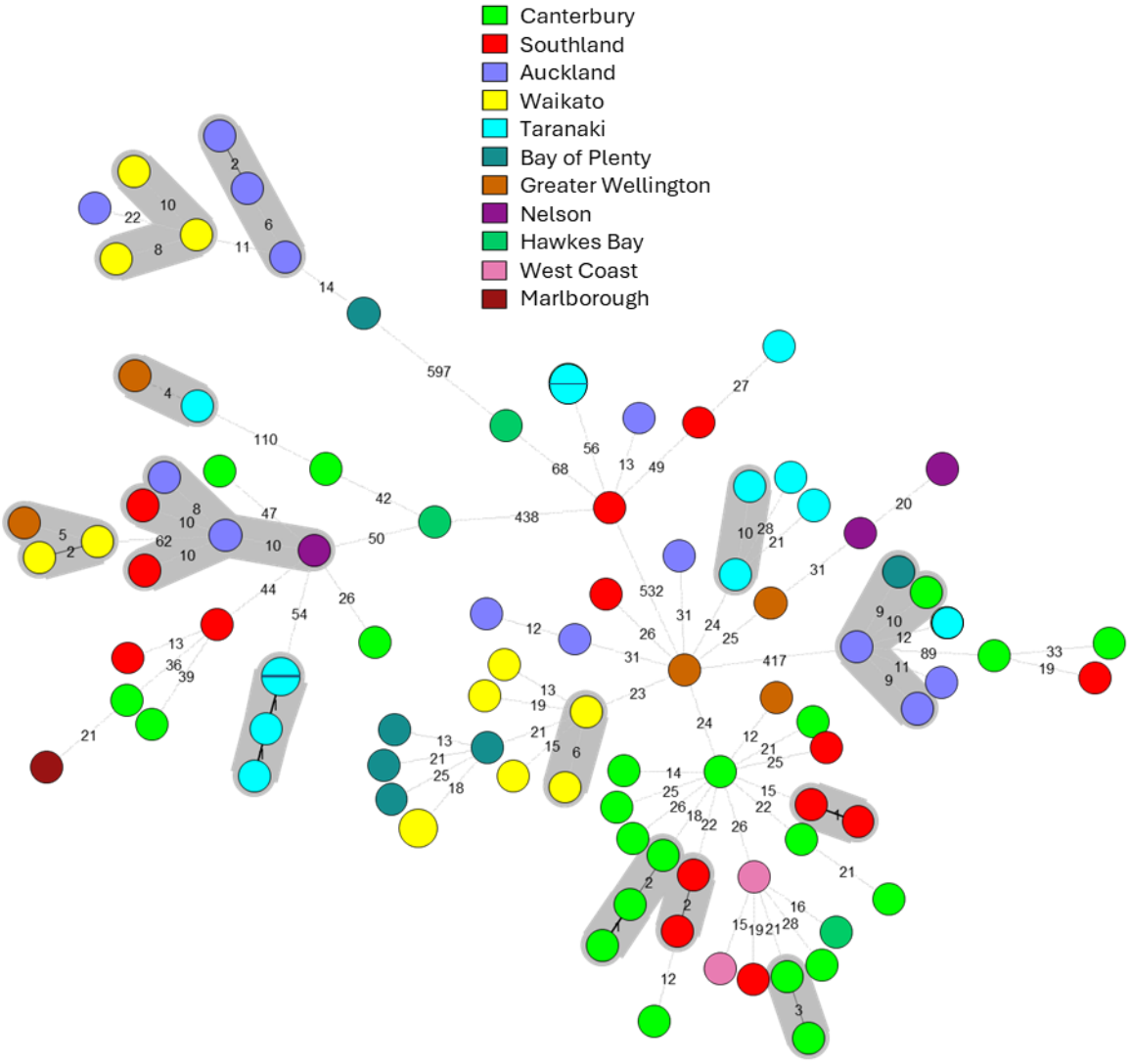
Seventeen different *Salmonella* serotypes were identified (Table 27). Typhimurium was the most frequently recovered serotype. All three Brandenburg isolates were recovered from the same site. Seven of the Give isolates were from one site with the other four from three other sites. The two Emek isolates were from the same site.

In contrast to *Campylobacter* cases, *Salmonella* case isolates are currently typed when it is possible to do so. All of the serotypes listed in Table 27 have been isolated from human clinical cases between 2020 and 2024, although some only infrequently (e.g., Ruiru, Emek, Derby). There were four different sequence types among the *S.* Typhimurium isolates (ST19, ST568, ST2297 and ST9767). Using wgMLST there was a significant degree of diversity within the *S.* Typhimurium (Figure 11), although several isolates with less than 10 wgMLST were from the same region. Similar to other *Salmonella* serotypes (Figure 12), there was a low level of diversity among isolates from the same region suggesting potential common sources.

Table 27: Frequency of serotypes of sequenced *Salmonella* in water samples.

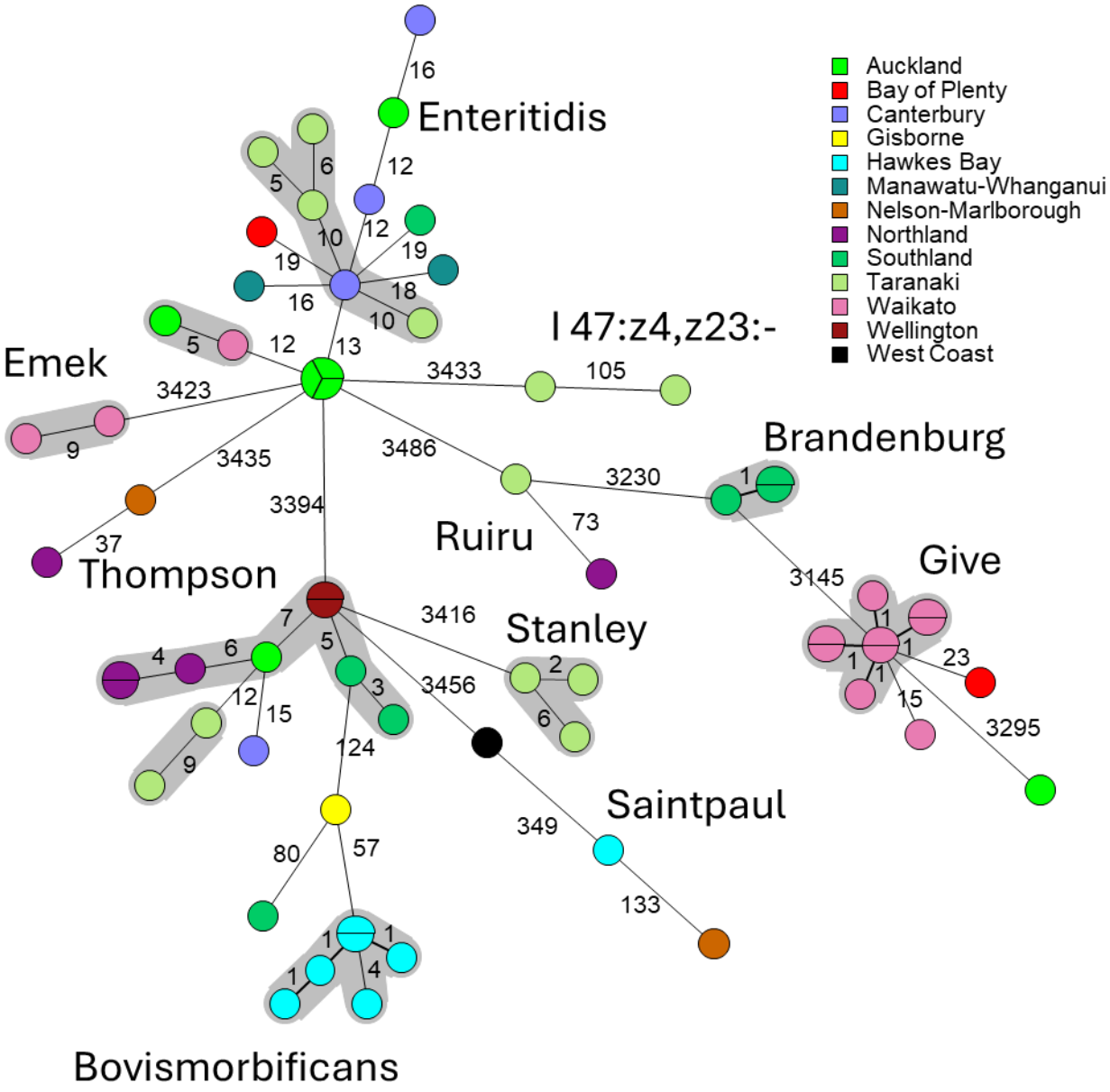
Serotype	Number of water samples
Typhimurium	94
Bovismorbificans	19
Enteritidis	17
Give	11
Brandenburg	3
Saintpaul	3
Stanley	3
Emek	2
I 47:z4,z23:-	2
Ruiru	2
Thompson	2
Agona	1
Derby	1
Kiambu	1
Livingstone	1
Mississippi	1
Schwarzengrund	1
Senftenberg	1
Total	**165**

Figure 11: Whole genome MLST minimum spanning tree of *Salmonella* Typhimurium



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences.

Figure 12: Whole genome MLST minimum spanning tree of non-Typhimurium *Salmonella* serotypes isolated more than once Taranaki



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different region of New Zealand. Grey shading indicates 10 or fewer wgMLST differences.

7.4 SHIGA TOXIN-PRODUCING *E. COLI*

There are at least six groups of pathogenic *E. coli* including enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enterohaemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EaggEC), and diffusely adherent *E. coli* (DAEC). The most frequently reported *E. coli* strains causing illness in New Zealand are shiga toxin-producing *E. coli* (STEC) (also known as verocytotoxin toxin-producing *E. coli* (VTEC)), which belong to the EHEC group.

Shiga toxin-producing *E. coli* are a group of approximately 200 different strains that cause illness by producing a toxin called “shiga toxin”, encoded by the *stx* gene. There are two distinct *stx* genes (*stx* 1 and *stx* 2), each producing different toxins (Wang et al, 2024). In addition to the toxin gene, they also need to adhere to the gastrointestinal tract which they do through a protein called intimin, encoded by the *eaeA* gene. However, a recent study has shown the *eaeA* gene was absent in 36% of isolates from notified cases of STEC infection in New Zealand between January 2016 and June 2023 (Horn et al, 2024).

There are several O-antigen serotypes of STEC. In 2024 there were 1,115 notified cases of STEC infection in New Zealand and serotyping data was available for 632 (57%) of these. The most commonly identified serotypes were *E. coli* O157:H7 (151 cases), O26:H11 (131 cases) and 128:H2 (68 cases). In 2024, 68 different O-antigen serotypes were associated with notified cases (Pattis I et al, 2025).

Methods for enumeration of STEC by culture analysis and qPCR are given in Appendix B. Strains were isolated from water samples using enrichment broth culture, followed by PCR directly from broth. Broths that were positive for either of the *stx* genes and *eaeA* were then plated, and 20 isolates tested again by PCR in pools of five. When a pool tested positive, individual isolates were then tested.

7.4.1 Isolation of shiga toxin-producing *E. coli* by culture

A total of 1,003 samples were assessed for the presence of STEC by enrichment culture. The *stx1*, *stx2*, *eaeA* and *hlyA* genes were frequently detected in the enrichment broths, but these were much less frequently identified in pools of isolates (Table 28). Of the 37 samples with either *stx1* and/or *stx2* detected in the pools of isolates, 30 had *stx2* only and 25 lacked *eaeA*.

Table 28: Frequency of detection of STEC genes from broths and pooled isolates

Gene Target	Number of water samples with STEC genes detected in broths (% of samples tested)			Water samples with gene detected in pooled isolates[a]		
	Detected	Weak	ND	Detected	Weak	ND
stx1	130 (13%)	63 (6%)	810 (81%)	5	2	327
stx2	166 (17%)	80 (8%)	757 (75%)	32	0	302
eaeA	916 (91%)	32 (3%)	55 (5%)	46	0	288
hlyA	136 (14%)	117 (12%)	750 (75%)	17	2	315

a: Isolate pools were taken from broths where *stx1* and/or *stx2* gene was present plus the *eaeA* gene.

Using the criterion of STEC being present if the *stx1* and/or *stx2* gene is detectable in the enrichment broth culture, there were 51 sites where STEC was detected (Table 29).

Table 29: Frequency of isolation of STEC in broths from different sites

Percentage of samples at a site with STEC detected	Number of sites	Samples tested	Samples with STEC present
81-100%	5	20	19
61-80%	4	38	26
41-60%	17	317	158
21-40%	18	260	77
1-20%	13	277	27
0%	14	91	0

7.4.2 Shiga toxin-producing *E. coli* qPCR analysis

qPCR analysis for the *stx1* and *stx2* genes was performed for 979 samples. For 856 of these (87%) neither gene was detected. The *stx1* gene was quantifiable in nine samples (2,300 GC/100 ml in one sample with the rest between 70 and 680 GC/100ml), and detectable but below the limit of quantification in a further 62 samples. The *stx*2 gene was quantifiable in 24 samples (30 – 520 GC/100 ml in all but two samples which had 2240 and 930 GC respectively), and detectable but below the limit of quantification in a further 61 samples. There were seven samples where both *stx1* and *stx2* were quantifiable, 10 samples where *stx2* was quantifiable and *stx1* was detectable but below the limit of quantification, and 17 samples were both genes were detectable but below the limit of quantification. These qPCR levels could reflect to up to six samples with >10 STEC / 100 ml, and another 5 samples with >1 STEC / 100 ml.

The direct qPCR analysis detected the *stx1* and/or *stx2* genes in less samples than the culture methods. Comparing the numbers of samples with weak or present culture results with the qPCR results revealed that qPCR only detected STEC in 23 – 30% of the samples where STEC was detected by culture methods (broth or pooled isolates). Conversely, qPCR detected STEC in 8% of samples where broth culture methods did not detect STEC and in 20% of pooled isolates where culture methods did not detect STEC.

Overall, the culture and qPCR results both suggest STEC was potentially present in up to 20 – 30% of the study samples, but the qPCR results indicate that concentrations in the water samples were low, with 97% of the samples below the limit of quantification.

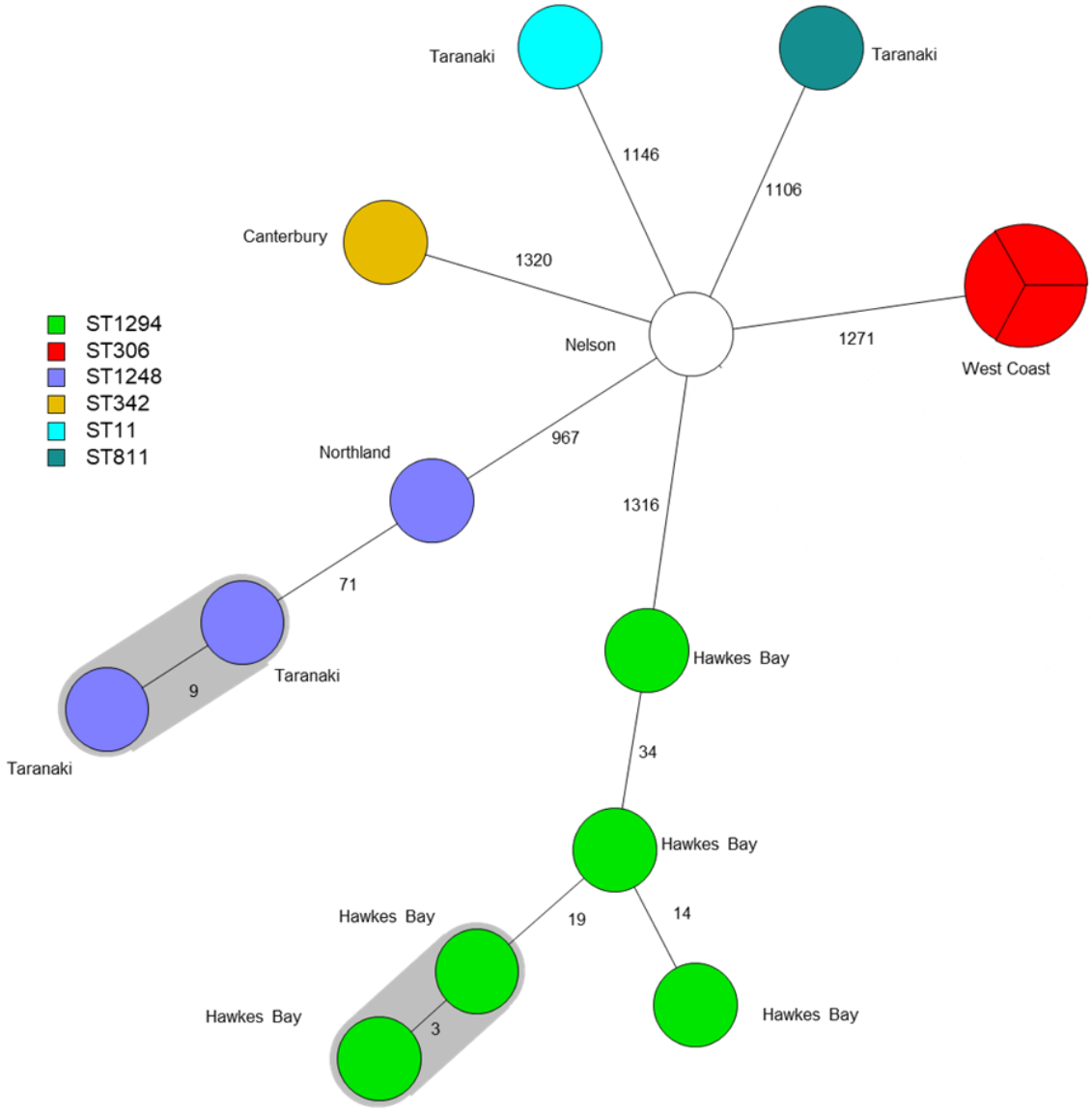
7.4.3 Whole genome sequencing of shiga toxin-producing *E. coli*

Isolates from 17 samples putatively identified as STEC based on the presence of *stx* genes were subjected to WGS. Four of these isolates were not *E. coli*, as shown in Table 30. Of the remaining isolates, three were classified as atypical EPEC (O-novel:H7, O84:H2 and O157:H8), and the rest all STECS. Four were O9a:H30, three O-novel:H21, and one each O9:H30, O128ac:H2 and O177:H25. These STEC have all been identified in clinical cases in New Zealand. Isolates which were more similar, were from the same region. (Figure 13).

Table 30: Summary of results from sequencing of putative STEC isolates

STEC qPCR results	Data from WGS			
	Predicted Type	Genes Detected	Serotype	ST
stx1*, *eaeA	atypical EPEC, STEC	stx1a,-det, eaeA, ehxA, rpoB-AEM.	O84:H2	ST306
stx2*, *eaeA	atypical EPEC, STEC	stx2c, stx2-det_F4_R1, eaeA, rpoB-AEM.	O-novel:H7	ST11
stx2*, *eaeA	atypical EPEC, STEC, UPEC	stx2c, eaeA, rpoB-AEM	O157:H8	
stx2	STEC	stx2B, ehxA, rpoB-AEM.	O128ac:H2	ST811
stx2*, *eaeA*, *hylA	STEC	*stx2c*, stx2-det_F4_R1, *eaeA*, *ehxA*, *rpoB*-AEM	O177:H25	ST342
stx2	STEC	stx2A,B, rpoB-AEM.	O9:H30	ST1294
stx2, eaeA	STEC	stx2A,B, rpoB-AEM.	O9a:H30	ST1294
stx2	STEC	stx2A,B, rpoB-AEM.	O9a:H30	ST1294
stx2	STEC	stx2A,B, rpoB-AEM.	O9a:H30	ST1294
stx2	STEC	stx2A,B, rpoB-AEM.	O9a:H30	ST1294
stx2	STEC	stx2A,B, rpoB-AEM.	O-novel: H21	ST1248
stx1	STEC	stx2A,B, rpoB-AEM.	O-novel:H21	ST1248
stx2	STEC	stx2A,B, rpoB-AEM.	O-novel:H21	ST1248
stx1	*Leclercia tamurae* (100%)			
Very weak *stx1*	*Providencia alcalifaciens* (75%), *Citrobacter portucalensis* (25%)			
Very weak *stx1*	*Kluyvera ascorbata* (100%)			
stx1*, *stx2	*Escherichia marmotae* (96%)			

Figure 13: Whole genome MLST analysis of STEC isolates



Numbers on branches are the number of wgMLST differences. Each colour corresponds to a different sequence type. Grey shading indicates 10 or fewer wgMLST differences.

7.5 PROTOZOA

Cryptosporidium and *Giardia* cause diarrhoea and abdominal pain after the ingestion of oocysts (*Cryptosporidium*) or cysts (*Giardia*) (hereon, (oo)cysts). (Oo)cysts may be present in the gut of humans and animals such as cattle, sheep, cats, dogs, rats and possums and can be transferred by infected faecal material. (Oo)cysts can survive adverse conditions in the environment for months until ingested by a new suitable host.

Several waterborne outbreaks of cryptosporidiosis and giardiasis have occurred in New Zealand and been attributed to contaminated drinking water (surface- and groundwater

sources), and exposure to recreational water and swimming pools (ESR - Public Health Surveillance: https://surv.esr.cri.nz/surveillance/surveillance.php).

Different protozoan species (e.g., *C. parvum* vs *C. hominis*) and sub/genotypes are known to have different host associations (Garcia–R et al, 2017).

7.5.1 *Cryptosporidium* and *Giardia* detection

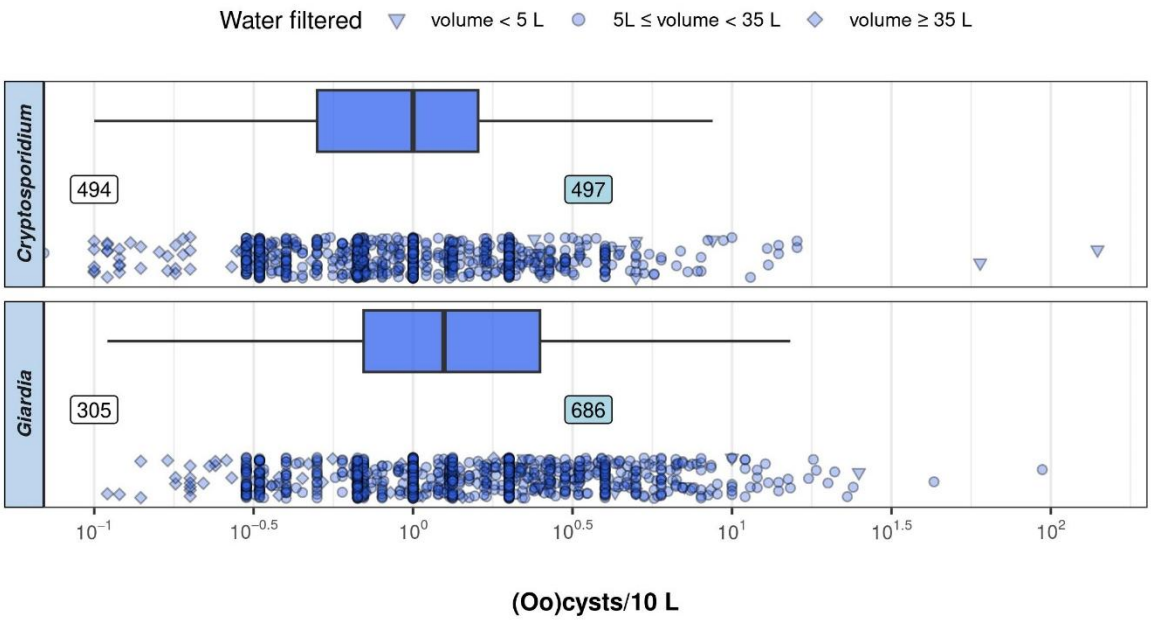
Cryptosporidium and *Giardia* were analysed in 991 samples using methods described in Appendix B. The volumes of water filtered ranged from 480 ml to 102 litres, with 74 samples having < 10 litres filtered, and 872 samples between 10 and 36 litres. Smaller volumes were filtered from samples with more turbid water which is often of lower water quality.

Cryptosporidium were detected in 497 samples (50%) and *Giardia* were detected in 686 samples (69%) (Table 31, Figure 14). A total of 377 samples had both *Cryptosporidium* and *Giardia* present.

Table 31: Summary of *Cryptosporidium* and *Giardia* concentrations in water samples

Organism	Samples	Samples greater than detection limit (%)	Recovered counts per 10 L					
			Min	25th %ile	50th %ile	75th %ile	95th %ile	Max
Cryptosporidium	991	497 (50%)	< 0.11	0.5	1.0	1.6	4	140
Giardia	991	686 (69%)	< 0.11	0.7	1.3	2.5	7.1	94

Figure 14: Distribution of *Cryptosporidium* and *Giardia* concentrations in water samples



Note: The numbers in boxes indicate the number of samples with concentrations below the limit of detection (white boxes), values within the range of quantification (blue boxes). Box plots show the range of values (whisker lines), boxes represent the 25th to 75th percentile range, and line in the box the median value.

While *Cryptosporidium* and *Giardia* were frequently detected in water samples, the recovered concentrations were generally low, with only one *Cryptosporidium* above and one *Giardia* close to a concentration of 100 per 10 litres, which is equivalent to 1 per 100 ml. Recovery experiments found between 20 – 56% of (oo)cysts in a water sample were recovered by laboratory testing during the Freshwater Survey (A Pita, Micro Aqua Tech Massey University, personal communication, August 2024). Applying a worst-case recovery rate of 20%, it is estimated that three and seven water samples had *Cryptosporidium* and *Giardia* concentrations above 1 per 100 ml respectively. For QMRA calculations a recovery rate of 20% will be used.

For most study sites, protozoa were repeatedly detected in water samples, as detailed in Table 32. Forty-seven out of 71 sites had *Cryptosporidium* detected in more than 40% of samples, and 45 out of 71 sites had *Giardia* detected in more than 60% of samples.

Table 32: Frequency of isolation of *Cryptosporidium* and *Giardia* at different sites

Percentage of samples at a site with *Cryptosporidium* detected	Number of sites	Samples tested	Samples with *Cryptosporidium* detected
81-100%	6[a]	34	30
61-80%	10	204	133
41-60%	31	495	257
21-40%	15	219	72
1-20%	4	32	5
0%	5	7	0
Percentage of samples at a site with *Giardia* detected	**Number of sites**	**Samples tested**	**Samples with *Giardia* detected**
81-100%	15[a]	172	153
61-80%	30	554	407
41-60%	13	212	116
21-40%	5	25	8
1-20%	2	16	2
0%	6	12	0

[a] Some sites only had one sample which was positive; 1 *Cryptosporidium*, and 3 *Giardia* 100% detected sites.

7.6 ENTERIC VIRUSES

There are a number of viruses that are potentially waterborne, including adenovirus, enterovirus, norovirus, polyomavirus, rotavirus, hepatitis A virus, hepatitis E virus, astrovirus and sapovirus (Sinclair et al, 2009). Norovirus, enterovirus and adenovirus were tested by qPCR during the Phase I of the study.

7.6.1 Adenovirus

Adenoviruses cause infections in all mammals and are very host-specific (e.g., bovine adenoviruses infect cows, human adenoviruses infect humans etc), and therefore, can be used for source tracking applications). Human adenoviruses are transmitted person to person most commonly via the faecal-oral route but also via aerosols contacting the mouth, nasopharynx or eyes. There are seven identified species (A-G) and over 50 types of human adenoviruses. Human adenovirus species F is a major cause of gastroenteritis in young children. Human adenoviruses are ubiquitous in the human population with most humans infected by age 20. Approximately half of the identified types are associated with human disease, but many infections are asymptomatic (Lynch and Kajon, 2016).

7.6.2 Enterovirus

Enteroviruses cause a wide range of diseases following exposure by the faecal-oral or airborne transmission routes. Asymptomatic/subclinical infections are common. Enteroviruses have been often used to assess human health risk due originally to their relative ease of culturing and more recently through detection by PCR. Enteroviruses have been used as a faecal indicator in a number of studies (Puig et al, 1994; Tani et al, 1995), with standard EPA methods developed (Cashdollar et al, 2013).

7.6.3 Norovirus

Gastroenteritis causing-norovirus genogroups I and II (GI and GII) are frequently responsible for disease outbreaks associated with faecally-contaminated water (Hewitt et al, 2007; Kukkula et al, 1997). They are excreted in large amounts in the faeces of infected individuals and can be present in high concentrations (> 10^3 genome copies/L) in municipal wastewater. As noroviruses are not persistently excreted, and with outbreaks often showing seasonal tendencies, their presence in wastewater may be more sporadic than human adenoviruses.

7.6.4 Virus analysis

Samples were only tested in phase 1. Full details of analysis are provided in Appendix B. To assess the presence of viruses in the water samples, 10 litre samples were concentrated using a modified hollow-fibre ultrafiltration followed by beef extract elution and polyethylene glycol 6,000 precipitation. Viral nucleic acid was extracted, and PCR / two-step RT-qPCR assays were used to detect adenovirus (Hernroth et al, 2002), enterovirus (Donaldson et al, 2002), norovirus genogroup I (Wolf et al, 2010), and norovirus GII (Kageyama et al, 2004).

Out of 52 samples, 21 samples contained potential RT-PCR inhibition of the viral analyses. A murine norovirus RT-qPCR control indicated that the recovery of viruses from water samples ranged between 3 – 12%. Table 33 shows the individual results of virus analysis for the 11 samples in which at least one target virus was detected in at least one replicate, a summary results for the other samples. All detections were below the limit of quantification. Many laboratories do not consider a sample as positive where only one of the three

replicates is positive, and/or where the CT[3] value is > 40. Using these criteria only five of the samples would be considered to have virus present, with norovirus GI in two samples and norovirus GII in three others (Green highlights in Table 33). However, even allowing for a more permissive criteria detection in at least one replicate then (a) None of the samples had detectable adenovirus; (b) Only two samples had enterovirus detected (both with only 1 of 3 replicates positive); and (c) Norovirus GI was potentially detected in another four samples, and norovirus GII in another five samples.

CrAssphage was detected in all samples in which a target virus was detected (Table 33), although three of the samples were below the limit of quantification. The samples with highest levels of CrAssphage, included four of the five samples with the strongest evidence of pathogenic virus presence.

Table 33: Viral qPCR results for samples tested in Phase I.

River (Samples)	HAdV CT	NoV GI CT	NoV GII CT	Enterovirus CT	CrAssphage	
					CT*	Copies / 100 ml
SI_10	ND/ND/ND	38.3/ND/39.1	ND/ND/ND	ND/ND/ND	26.6	337,000
NI_37	ND/ND/ND	ND/ND/38.8	38.2/ND/38.5	ND/39.9/ND	23.9	314,200
NI_37	ND/ND/ND	ND/ND/ND	38.2/38.3/37.4	ND/ND/ND	25.6	229,777
SI_10	ND/ND/ND	39.7/38.4/39.9	ND/ND/39.3	ND/ND/ND	28.8	67,100
NI_37	ND/ND/ND	ND/ND/ND	ND/ND/39.7	ND/ND/ND	28.3	51,200
8 rivers (20)	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND/ND/ND	30.8-36.7	24,600 - 655
NI_43	ND/ND/ND	37.9/ND/ND	40.1/ND/ND	ND/ND/ND	37.7	537
NI_43	ND/ND/ND	41.3/40.5/ND	ND/ND/ND	ND/ND/ND	35.2	475
3 rivers (3)	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND/ND/ND	35.4-36.2	420 - 257
NI_43	ND/ND/ND	ND/ND/ND	ND/39.7/ND	ND/ND/ND	35.6	224
NI_34 (2)	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND/ND/ND	36.5 36.6	149 99
SI_12	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND/39.2/ND	37.7	DBLQ
NI_8	ND/ND/ND	ND/40.5/38.9	ND/ND/ND	ND/ND/ND	37.8	DBLQ
NI_2	ND/ND/ND	ND/ND/ND	37.8/37.1/ND	ND/ND/ND	38.1	DBLQ
7 rivers (15)	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND/ND/ND	ND	ND

Notes: CT, cycle threshold for each of three replicates; CT* for CrAssphage all had 2 replicates, with average provided. HAdV, human adenovirus; NoV GI, norovirus GI; NoV GII, norovirus GII; DBLQ – detected below limit of quantitation. ND, not detected. The copies/100 ml measurement depends on the volume of water filtered which varied between samples.

[3] CT (cycle threshold) is the cycle of PCR when amplification is first detected, with the lower the CT, the more copies of the target present.

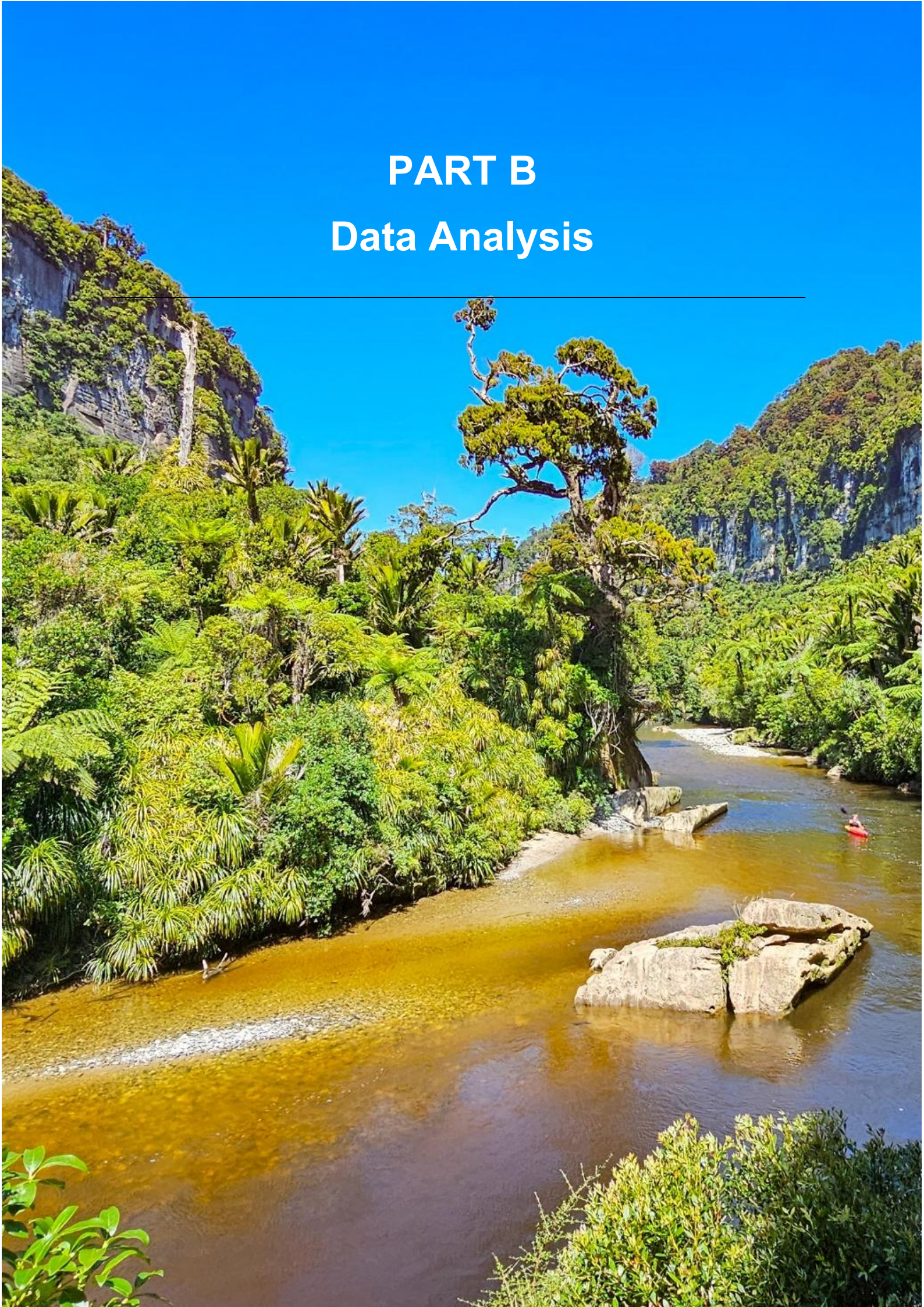
7.6.5 Virus testing in phases 2 and 3

The decision was made to not test for viral pathogens in phases 2 and 3. This was based on the following considerations.

1. Despite selecting sites with high levels of contamination, and based on MST results, with human faecal contamination likely in at least half of the samples, no adenovirus were detected, and enterovirus and norovirus detections were at the limit of detection and with low reproducibility.
2. None of the target virus detections were above the limit of quantification, meaning only presence/absence data was likely to be generated.
3. These observations may reflect methodological challenges rather than an absence of pathogenic virus at levels of concern.
4. CrAssphage as a viral indicator provides a more robust measure of potential viral presence and can be used to estimate the likely presence of viruses based on the relative levels of crAssphage to pathogenic viruses.
5. Pathogenic virus analysis was the most expensive assay and as such removing this analysis significantly reduced both courier and laboratory analysis costs. The laboratory which performed this testing was also heavily involved in wastewater testing during the COVID pandemic.

PART B

Data Analysis



8 PREDICTING THE PRESENCE OF PATHOGENS

Key findings

- *The probability of the presence of Campylobacter or Salmonella in a water sample increases with increasing concentrations of E. coli, enterococci, human, ruminant or wild bird MST markers.*
- *The probability of the presence of Campylobacter or Salmonella in a water sample also increased with the number of sources identified.*
- *The probability of the presence of STEC in a water sample increases with increasing concentrations of E. coli, enterococci or the ruminant MST marker. While STEC has been identified in some water samples with human and wild bird markers, there is no evidence for STEC presence being related to the concentration of human or wild bird MST markers.*
- *Neither E. coli or enterococci concentrations are predictors for the probability of the presence of Cryptosporidium or Giardia in the water samples.*

8.1 INTRODUCTION

The 2020-2024 survey collected presence and absence data for five pathogens: *Campylobacter* (*C. jejuni* and *C. coli*), *Salmonella*, STEC, *Cryptosporidium*, and *Giardia*. This section investigates the presence or absence of these pathogens alongside the concentration of two faecal indicator bacteria (*E. coli* and enterococci) and five microbial source tracking markers (human-associated: crAssphage, HF183, BiADO; ruminant: BacR, and wild bird, GFD) collectively referred to as indicators. The final part of this section compares the sequence types of *C. jejuni* that were identified in the survey with both known sources of these sequence types and the MST profiles of the water samples.

Rates of detection between pathogens cannot be directly compared when different volumes of water were analysed. The larger the volume of water tested, the more likely a pathogen will be detected if present at the sampling site. In Phase I of the project, *Campylobacter*, *Salmonella* and STEC MPN had a total tested water volume of 903 ml per sample. In Phase II and III, *Campylobacter* had a total tested water volume of 1030 ml and Salmonella and STEC were tested for presence or absence in 1 L. *Cryptosporidium* and *Giardia* between one and 100 L, but mostly greater than 10 L volumes of water were tested per sample (section 7.5.1).

8.2 ANALYSIS APPROACH

Before analysis of the data, the following data cleaning steps were undertaken:

- Removal of data relating to water samples where the salinity was too high to be classed as freshwater.
- Removal of data relating to sites where fewer than 5 samples were collected because the statistical analysis included site as a factor. Leaving 56 sites.
- Removal of data with missing data in the pathogen-indicator combination.

Two approaches have been taken; (i) statistical modelling of univariate relationship between a single pathogen and a single indicator and (ii) considering the frequency of detection of pathogens for different MST profiles.

The relationship between the pathogens and the indicators was explored using a binomial generalised linear mixed model (GLMM). A binomial model is used to represent the two distinct outcomes of the pathogen being present or not.

The probability that a water sample will have the pathogen $\mathbf{X}$ present, $\mathbf{p(X)}$, is modelled by the relationship in equation 8.1.

$$\boldsymbol{ln}\left(\frac{\boldsymbol{p(X)}}{\boldsymbol{1-p(X)}}\right) = \boldsymbol{Intercept} + (\boldsymbol{\beta} \times \boldsymbol{Indicator\ concentration}) + \boldsymbol{f(Site)} + \varepsilon \qquad [8.1]$$

The left-hand side of equation 8.1 uses a transformation to allow the probability of a pathogen being present to be represented by a linear combination of factors on the right-hand side of the equation. The intercept and $\boldsymbol{\beta}$ are constants derived during model fitting.

On the right-hand side of the equation, the model includes a fixed effect for the indicator concentration, meaning the same relationship will apply across all the sites. However, the variability in pathogen presence across different sites and repeated sampling at the same site are included in the model by adding a random effect for site, $\boldsymbol{f(Site)}$. The modelling results, not shown, suggest there are differences between sites. The last part of the right hand side of equation 8.1 is the residual (ε) not explained by the model.

The results for the model are presented in terms of odds ratios. Odds are the probability that a pathogen will be present divided by the probability that a pathogen will not be present:

$$\boldsymbol{odds} = \frac{\boldsymbol{p(X)}}{\boldsymbol{1-p(X)}} \ . \qquad [8.2]$$

The odds ratios presented in this section are the change in odds due to a change of one $\log_{10}$ value in the indicator concentration. An odds ratio of one means there is no change in odds with changes in the concentration of the indicator after accounting for site. Odds ratios greater than one indicate an increase in the odds of the pathogen being present with increasing indicator concentration and conversely an odds ratio of less than one indicates a decrease in odds of the pathogen being present with increasing indicator concentration.

Note the indicators have different scales that have not been standardised for this analysis, so the odds ratios for different indicators should not be compared, but the odds ratios can be compared for the same indicator with different pathogens.

The models do not fully explain the variability in the pathogen presence given the site and indicator factors. The uncertainty in the odds ratio are indicated by the 95% confidence

intervals provided in the results. Another model output relating to the model fit is the *p* value. The *p* value gives the probability of obtaining the observed odds ratio if the presence of the pathogen at freshwater sites was not a function of the indicator concentration in the water (β=0 in Equation 8.1).

As typical examples of the data, Figure 15 and Figure 16 show the detection of pathogenic *Campylobacter* (*C. jejuni* and/or *C. coli*) or STEC in water samples with different GFD concentrations (red dots at a probability of one indicate the pathogen is detected, and green dots at a probability of zero the pathogen is not detected). There is no distinct boundary in the indicator concentration when the pathogen is likely to be present or not. The two box plots comparing the GFD concentrations separated by the presence or absence of STEC show little difference in the distributions of GFD, and the upper plot showing the estimated probability of detecting STEC is close to flat (odds ratio close to one).

In contrast, the plots for *Campylobacter* and GFD show overlapping but different distributions for GFD concentrations for water samples where *Campylobacter* was present or not. The plot of estimated probability of detecting *Campylobacter* changes from 0.2 to 0.95 (odds ratio of 5.3) over the range of observed GFD concentrations.

Figure 15: The presence/absence and probability of detection of STEC by GFD concentration.

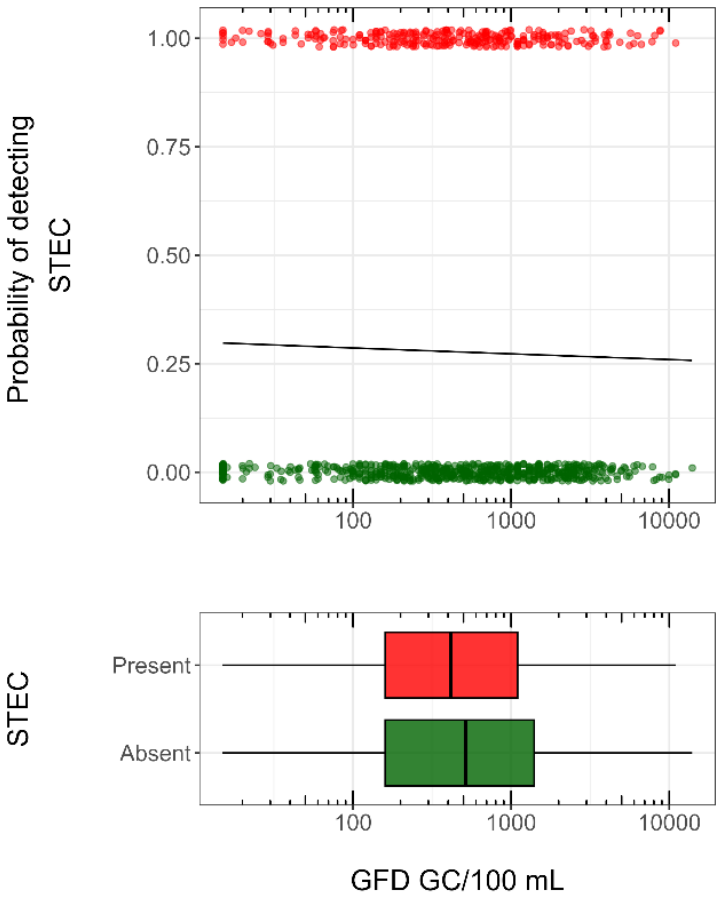
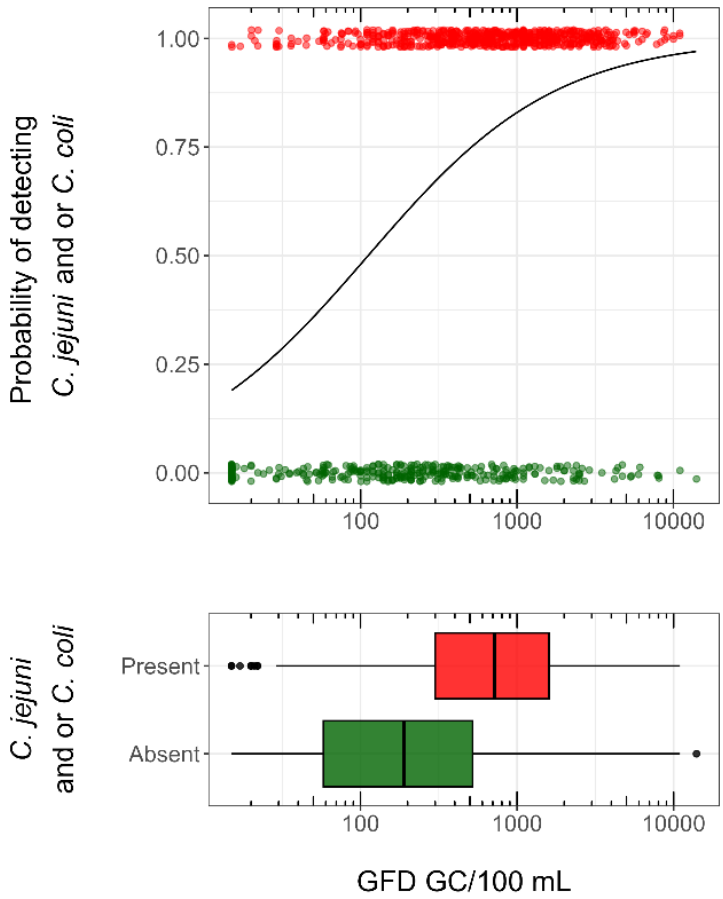


Figure 16: The presence/absence and probability of detection of *Campylobacter* by GFD concentration.



8.3 PRESENCE OF *CAMPYLOBACTER*

The results from modelling the probability of pathogenic *Campylobacter* presence as dependent on concentrations of each indicator are shown in Table 34. For all indicators, inclusive of FIB and MST markers, the confidence intervals for the odds ratios were above a value of one and p-values < 0.001, providing evidence for an increasing trend of probability of pathogen presence with indicator concentration.

The frequency of detection of *C. jejuni* at different concentrations for water samples contaminated with varying combinations of faecal sources is shown in Table 35. The highest detection rates of *C. jejuni* were in samples with more than one marker detected. These samples were also more likely to have higher concentrations of *C. jejuni*.

Samples with no faecal source identified, or only low MST markers had the lowest percentage of samples where *C. jejuni* was isolated (38%), compared with samples with a high level of at least one MST marker (71%).

Table 34: Odds ratios for the presence of *C. jejuni* and or *C. coli* in water samples as predicted by log_{10} scaled indicators using the univariate GLMM model.

Indicator	Indicator type	Odds ratio	95% confidence interval for odds ratio	*p* value
E. coli	Faecal matter	2.8	2.1 - 3.8	**< 0.001**
Enterococci	Faecal matter	4.2	3.1 - 5.8	**< 0.001**
crAssphage	Human source	2.0	1.4 - 2.7	**< 0.001**
HF183	Human source	2.2	1.7 - 3.0	**< 0.001**
BiADO	Human source	2.1	1.5 - 2.9	**< 0.001**
BacR	Ruminant source	1.9	1.6 - 2.4	**< 0.001**
GFD	Wild bird source	5.3	3.6 - 7.8	**< 0.001**

Table 35: *C. jejuni* isolation from freshwater samples with different MST profiles. Ranked from highest to lowest detection of *C. jejuni*.

MST Source	*C. jejuni* MPN/100 ml. Number of samples (%)					Total
	ND	0.2 – 0.7	3 – 9.2	14 to 93	>100	
Human, ruminant and wild bird	12 (15%)	9 (12%)	18 (23%)	30 (38%)	9 (12%)	78
Ruminant and wild bird	35 (16%)	38 (17%)	56 (25%)	80 (36%)	16 (7%)	225
Low ruminant and wild bird	9 (19%)	10 (21%)	16 (34%)	10 (21%)	2 (4%)	47
Human and wild bird	26 (25%)	26 (25%)	31 (30%)	18 (17%)	4 (4%)	105
Low ruminant and low wild bird	13 (36%)	9 (25%)	8 (22%)	6 (17%)	-	36
Wild bird	38 (39%)	34 (35%)	13 (13%)	10 (10%)	3 (3%)	98
Ruminant	84 (49%)	43 (25%)	29 (17%)	14 (8%)	3 (2%)	173
Human	24 (51%)	15 (32%)	4 (9%)	3 (6%)	1 (2%)	47
Low wild bird	20 (57%)	8 (23%)	5 (14%)	2 (6%)	-	35
Aged human	14 (58%)	3 (13%)	1 (4%)	4 (17%)	2 (8%)	24
Low ruminant	29 (67%)	7 (16%)	6 (14%)	1 (2%)	-	43
No source identified	48 (77%)	9 (15%)	2 (3%)	2 (3%)	1 (2%)	62
Total	**355 (36%)**	**213 (22%)**	**192 (19%)**	**183 (19%)**	**42 (4%)**	**985**

8.4 PRESENCE OF *SALMONELLA*

In phase 2 and 3, only presence/absence testing in 1 L of water was conducted. The results from modelling the probability of *Salmonella* presence as dependent on concentrations of each indicator are shown in Table 36. Similar to *Campylobacter*, for all indicators, the confidence intervals for the odds ratios were above a value of one and p-values < 0.001, providing evidence for an increasing trend of probability of *Salmonella* presence with indicator concentration.

Table 36: Odds ratios for the presence of *Salmonella* in water samples as predicted by log_{10} scaled indicators using univariate GLMM.

Indicator	Indicator type	Odds ratio	95% confidence interval for odds ratio	*p* value
E. coli	Faecal matter	3.1	2.3 - 4.1	**< 0.001**
Enterococci	Faecal matter	2.8	2.2 - 3.7	**< 0.001**
crAssphage	Human source	2.1	1.6 - 2.7	**< 0.001**
HF183	Human source	2.1	1.7 - 2.7	**< 0.001**
BiADO	Human source	2.2	1.7 - 2.9	**< 0.001**
BacR	Ruminant source	1.7	1.4 - 2.1	**< 0.001**
GFD	Wild bird source	2.2	1.5 - 3.2	**< 0.001**

Salmonella were more frequently isolated using culture from samples that included human faecal inputs (32%), than samples with any other MST profile (4 – 19%) (Table 37). There was again a cumulative effect with samples with more than one source, more likely to have *Salmonella* detected.

Table 37: *Salmonella* presence by culture in samples with different MST profiles.

MST Category	Present	Total Tested
Human, ruminant and wild bird	38 (49%)	78
Human and wild bird	27 (26%)	105
Aged human	6 (25%)	24
Human	10 (21%)	47
Low ruminant and wild bird	7 (15%)	47
Wild bird	12 (12%)	98
Ruminant	21 (12%)	173
Ruminant and wild bird	27 (12%)	225
Low ruminant and low wild bird	4 (11%)	36
No source identified	4 (6%)	62
Low wild bird	2 (6%)	35
Low ruminant	-	43
Total	**157 (16%)**	**961**

8.5 PRESENCE OF STEC

The results from modelling the probability of STEC presence, based on PCR detection of stx genes in broth culture, as dependent on concentrations of each indicator are shown in Table 38.

For the faecal indicator bacteria, *E. coli* and enterococci, the confidence intervals for the odds ratios were above a value of one and *p*-values < 0.001. Providing evidence for an increasing trend of probability of STEC presence with increasing FIB indicator concentration, but at a lower odds ratio than that estimated for *Campylobacter* or *Salmonella*.

Of the MST markers, only the ruminant marker (BacR) is providing evidence for an increasing trend of the probability of STEC presence with increasing indicator concentration. There is no evidence for a relationship between the human or wild bird marker concentrations and the presence of STEC.

When samples were enriched by culture for STEC, samples with ruminant MST markers were more likely to have *Stx1* and/or *Stx2* gene detected (42%), than samples without ruminant MST markers (12%) (Table 39). Of the 13 water samples from which STEC were isolated, all had ruminant MST markers identified – nine with ruminant markers only; two with human and ruminant markers; and four with low level ruminant markers.

These analyses suggest that while STEC could be in rivers from human and wild bird sources, the dominant source is likely to be from ruminants. This is consistent with previous New Zealand studies of ruminants (Cookson et al. 2006).

Table 38: Odds ratios for the presence of STEC in water samples as predicted by log_{10} scaled indicators using univariate GLMM.

Indicator	Indicator type	Odds ratio	95% confidence interval for odds ratio	*p* value
E. coli	Faecal matter	2.1	1.6 - 2.7	**< 0.001**
Enterococci	Faecal matter	1.5	1.2 - 1.9	**< 0.001**
crAssphage	Human source	1.2	0.9 - 1.6	0.283
HF183	Human source	1.1	0.9 - 1.4	0.544
BiADO	Human source	0.9	0.7 - 1.2	0.629
BacR	Ruminant source	2.2	1.8 - 2.5	**< 0.001**
GFD	Wild bird source	0.9	0.7 - 1.3	0.69

Table 39: Samples with shiga toxin genes detected in broth culture compared with MST source.

MST Category	Present	Total Tested
Ruminant	87 (53%)	165
Ruminant and wild bird	98 (45%)	216
Low ruminant and low wild bird	14 (39%)	36
Human, ruminant and wild bird	26 (34%)	77
Aged human	8 (33%)	24
Low ruminant and wild bird	11 (23%)	47
Low ruminant	9 (21%)	43
No source identified	9 (15%)	61
Human and wild bird	14 (14%)	102
Human	3 (6%)	47
Wild bird	6 (6%)	98
Low wild bird	2 (6%)	33
Total	**281 (30%)**	**936**

8.6 PRESENCE OF *CRYPTOSPORIDIUM* AND *GIARDIA*

The results from modelling the probability of *Cryptosporidium* or *Giardia* presence as dependent on concentrations of each indicator are shown in Table 40 and Table 41. For all indicators and *Cryptosporidium,* the confidence intervals for the odds ratios were above but close to a value of one suggesting only a weak trend in increasing probability of *Cryptosporidium* presence with increasing indicator concentration. The evidence for the trend was particularly weak for human and wild bird markers ($p > 0.05$).

Table 40: Odds ratios for the presence of *Cryptosporidium* in water samples as predicted by log_{10} scaled indicators using univariate GLMM.

Indicator	Indicator type	Odds ratio	95% confidence interval for odds ratio	*p* value
E. coli	Faecal matter	1.3	1.1 - 1.6	0.005
Enterococci	Faecal matter	1.3	1.0 - 1.5	0.020
crAssphage	Human source	1.1	1.0 - 1.3	0.112
HF183	Human source	1.1	1.0 - 1.4	0.054
BiADO	Human source	1.1	0.9 - 1.3	0.225
BacR	Ruminant source	1.2	1.1 - 1.4	0.004
GFD	Wild bird source	1.3	1.0 - 1.6	0.068

The *Giardia* data also showed a weak trend of increasing probability of *Giardia* concentration with the indicators (Table 41), with the strongest associations with the wild bird marker and HF183 (which can be found in possum faeces in addition to human faeces).

Table 41: Odds ratios for the presence of *Giardia* in water samples as predicted by log_{10} scaled indicators using univariate GLMM.

Indicator	Indicator type	Odds ratio	95% confidence interval for odds ratio	*p* value
E. coli	Faecal matter	1.4	1.1 - 1.7	0.002
Enterococci	Faecal matter	1.2	1.0 - 1.5	0.051
crAssphage	Human source	1.3	1.1 - 1.6	0.006
HF183	Human source	1.5	1.3 - 1.9	<0.001
BiADO	Human source	1.3	1.1 - 1.6	0.006
BacR	Ruminant source	1.2	1.0 - 1.3	0.053
GFD	Wild bird source	1.8	1.4 - 2.3	<0.001

Cryptosporidium prevalence did not differ significantly in samples with different MST sources (Table 42). *Giardia* were found in samples with all MST profiles, with the only difference being that *Giardia* were more frequently identified in samples that included human and/or wild bird markers.

Table 42: Samples with *Cryptosporidium* or *Giardia* detected compared with MST source.

MST Category	*Cryptosporidium*	*Giardia*	Total Tested
Human and wild bird	51 (50%)	83 (82%)	101
Human, ruminant and wild bird	36 (52%)	54 (78%)	69
Low ruminant and low wild bird	14 (39%)	28 (78%)	36
Human	27 (59%)	35 (76%)	46
Ruminant and wild bird	115 (53%)	159 (74%)	215
Wild bird	49 (51%)	71 (73%)	97
Low wild bird	13 (39%)	23 (70%)	33
Other	6 (46%)	9 (69%)	13
Ruminant	93 (54%)	110 (64%)	171
Low ruminant and wild bird	23 (50%)	29 (63%)	46
Aged human	8 (33%)	13 (54%)	24
Low ruminant	19 (48%)	20 (50%)	40
No source identified	25 (42%)	30 (50%)	60
Total	**473 (50%)**	**655 (70%)**	**938**

8.7 C. JEJUNI WHOLE GENOME SEQUENCING AND MICROBIAL SOURCE TRACKING

For 494 water samples both MST and WGS analysis of *C. jejuni* was undertaken. As noted previously 46% of the sequenced *C. jejuni* isolates are of an ST that has previously been isolated from a human, poultry or ruminant host. Isolates from these sources have all been demonstrated to cause disease in humans, and therefore, pose a health risk.

When samples with different MST profiles are ranked in terms of the proportion of samples with a confirmed human infective ST, samples with human MST markers have higher proportion of human infective ST (Table 42). This suggests that at least some of the *Campylobacter* in these samples are coming from human faecal sources.

Samples with only ruminant or only wild bird MST markers also had ST likely to infect humans, again supporting these as sources of human infective *Campylobacter*.

Novel ST were present in higher proportions in samples with only low levels of MST markers, potentially indicating wildlife or less common avian species as possible sources for some of these. There was no difference in the quantitative levels of *C. jejuni* according to MLST type.

Table 42. Ranked table of the proportion of *C. jejuni* STs previously isolated from human, ruminant or poultry sources in samples with MST profiles.

MST Source	Number (%) of *C. jejuni* isolates with an ST previously isolated from a faecal source or novel ST			
	Human, ruminant or poultry	Wild bird	Novel	Total
Aged human	7 (88%)	1 (12%)	0	8
Human	10 (67%)	5 (33%)	0	15
Human and wild bird	32 (52%)	26 (42%)	4 (6%)	62
Human, ruminant and wild bird	28 (50%)	18 (32%)	10 (18%)	56
Ruminant	33 (47%)	26 (37%)	11 (16%)	70
Ruminant and wild bird	73 (47%)	52 (33%)	31 (20%)	156
Wild bird	17 (44%)	14 (36%)	8 (21%)	39
Low ruminant and wild bird	11 (41%)	8 (30%)	8 (30%)	27
Low ruminant &/or low wild bird	12 (32%)	14 (37%)	12 (32%)	38
No source identified	2 (13%)	6 (40%)	7 (47%)	15
Total	**225 (46%)**	**170 (35%)**	**91 (19%)**	**486**

8.8 FIELD VARIABLES AND BACTERIAL PATHOGEN PRESENCE

Field data collected at the time of water sampling (appendix D) allowed the same approach as described above (section 8.2) to see if the presence of *C. jejuni* or *C. coli*, *Salmonella* and STEC at concentrations as detected by this study (~1 per litre), changed with values of the collected field data. The purpose of this analysis was to determine which field variables may

be important for future modelling and to see if there were differences in the results for the different bacteria.

The rainfall and turbidity (Formazin nephelometric units, FNU) data each produced a very skewed distribution with a few much higher values which may bias the outcomes to a few values. Rather than apply a transformation which may make the interpretation of the odds ratios less intuitive, the data for rainfall and turbidity was stratified. Rainfall stratification used break points at 0, 2, 10, 20, 50, 75,100 and above 100 mm in a 24 hour or 72 hour period prior to sampling. Turbidity stratification used break points at 10, 50, 100, 500 and above 500 FNU.

This approach is sensible given the purpose of the analysis but would need to be further explored if the purpose changed to having a predictive purpose. For rainfall and turbidity, the odds ratio then becomes the change in odds from moving from one interval to the next highest one. For pH, water temperature and dissolved oxygen, the odds ratio is the change in odds due to a single unit value of the variable.

As for the analyses in the previous sections, data was used from sites for which 5 or more water samples were available with no missing data for the variables being considered. Differences between sites and repeated sampling from the same site was included in the model structure.

The results from modelling the presence of *C. jejuni* or *C. coli*, *Salmonella* and STEC as dependent on field variables are shown in Table 43. In summary:

- Increasing rainfall and turbidity resulted in an increasing odds of detection of *C. jejuni* or *C. coli*, *Salmonella* and STEC.
- Cumulative rainfall from 0 - 24 hrs prior to sampling resulted in a slightly higher odds ratio compared to cumulative rainfall from 0 - 72 hours prior to sampling for *Campylobacter* and *Salmonella* but was the same for STEC.
- For STEC, only rainfall and turbidity out of the measured field variables indicated a change in the odds of STEC being detected with a change in the level of the field variable.
- *C. jejuni* or *C. coli* was detected in the water samples with reduced odds when the values of conductivity (log_{10} transformed), dissolved oxygen, pH and water temperature increased.
- *Salmonella* was detected in water samples with reduced odds as the pH of the water increased.

Table 43: Odds ratios for the presence of bacterial pathogens in water samples as predicted by environmental field variables using a univariate GLMM model (ordered by odds ratio for each pathogen).

	Field variable	Odds ratio	95% confidence interval for odds ratio	*p* value[c]	Number of sites	Number of water samples
C. jejuni or C. coli	Conductivity (log_{10} µS/cm)	0.18	0.04 – 0.73	**0.016**	52	666
	pH	0.43	0.29 – 0.63	**<0.001**	48	700
	Dissolved oxygen (mg/L)	0.83	0.74 – 0.94	**0.004**	56	946
	Water temperature	0.85	0.80 – 0.90	**< 0.001**	56	972
	Rainfall (mm) intervals[a] (0 - 24 h before sampling)	1.36	1.15 – 1.61	**<0.001**	45	797
	Rainfall (mm) intervals[a] (0 - 72 h before sampling)	1.15	1.03 – 1.28	**0.015**	45	797
	Turbidity FNU intervals[b]	1.61	1.07 – 2.42	**0.02**	35	405
Salmonella	Conductivity (log_{10} µS/cm)	0.75	0.28 – 2.00	0.56	52	665
	pH	0.47	0.31 – 0.71	**<0.001**	48	699
	Dissolved oxygen (mg/L)	0.89	0.78 – 1.00	0.05	56	945
	Water temperature	0.98	0.93 – 1.04	0.56	56	971
	Rainfall (mm) intervals[a] (0 - 24 h before sampling)	1.49	1.29 – 1.72	**<0.001**	45	796
	Rainfall (mm) intervals[a] (0 - 72 h before sampling)	1.33	1.20 – 1.48	**< 0.001**	45	796
	Turbidity FNU intervals[b]	2.23	1.50 – 3.33	**<0.001**	35	405
STEC	Conductivity (log_{10} µS/cm)	0.43	0.15 – 1.24	0.12	52	650
	pH	0.86	0.61 – 1.21	0.39	48	683
	Dissolved oxygen (mg/L)	0.96	0.86 – 1.08	0.50	56	923
	Water temperature	1	0.95 – 1.05	0.93	56	949
	Rainfall (mm) intervals[a] (0 - 24 h before sampling)	1.31	1.15 – 1.49	**<0.001**	45	777
	Rainfall (mm) intervals[a] (0 - 72 h before sampling)	1.31	1.19 – 1.43	**< 0.001**	45	777
	Turbidity FNU intervals[b]	1.77	1.22 – 2.57	**0.002**	35	395

Notes:

a: Cumulative 24 hour rainfall data split into seven ordered intervals using break points of 0,2,10,20,50,75,100 and > 100 ml rain.

b: Turbidity data split into five ordered intervals using break points of 0, 10, 50, 100, 500 and >500 FNU.

c: Significance was characterised as $p < 0.05$ for all statistical analyses.

These results should be considered in the context of the range of field variable values recorded from the water samples (conductivity up to 1,000 µS/cm, dissolved oxygen up to 20.5 mg/L, pH 4.5 to 9.9 and water temperature 5 to 27°C) and not be extrapolated to general statements. Note also that water temperature showed a clear seasonal pattern (Appendix D, Figure 40), therefore, changes in odds ratios due to increases in water temperature may also reflect other seasonal changes such as UV levels and durations, or the presence of seasonal activities in the catchment.

While rainfall and turbidity both indicated an increased odds of pathogen presence for all three bacteria considered, there was high between-site variability. This suggests that these are not suitable variables to be applied to a national guideline value. However, field variables should be considered when developing site specific QMRAs.

9 PATHOGEN CONCENTRATIONS

Key findings

- *The concentration of Campylobacter, Cryptosporidium and Giardia in water samples is highly variable for given E. coli or enterococci concentrations.*
- *The Campylobacter concentration distribution's shape and range of values changes with E. coli and enterococci concentration in the water sample.*
- *The median concentration of Campylobacter increases with E. coli and enterococci concentrations up to about 1000 E. coli /100 ml after which there is less evidence of an increase.*
- *The Campylobacter concentrations are slightly lower at lower E. coli or enterococci concentrations using data from within the bathing season period (defined as 31 October to 31 March), compared to water samples collected outside this period.*
- *Most Cryptosporidium and Giardia concentrations are below 10 cysts or oocysts per 10 L. With such low concentrations, the Cryptosporidium or Giardia distributions minimally change with E. coli or enterococci concentration.*

The study provided concentration data for three pathogens: *Campylobacter (C. jejuni and C. coli)*, *Cryptosporidium* and *Giardia*. This section compares the pathogen concentration data with FIB data.

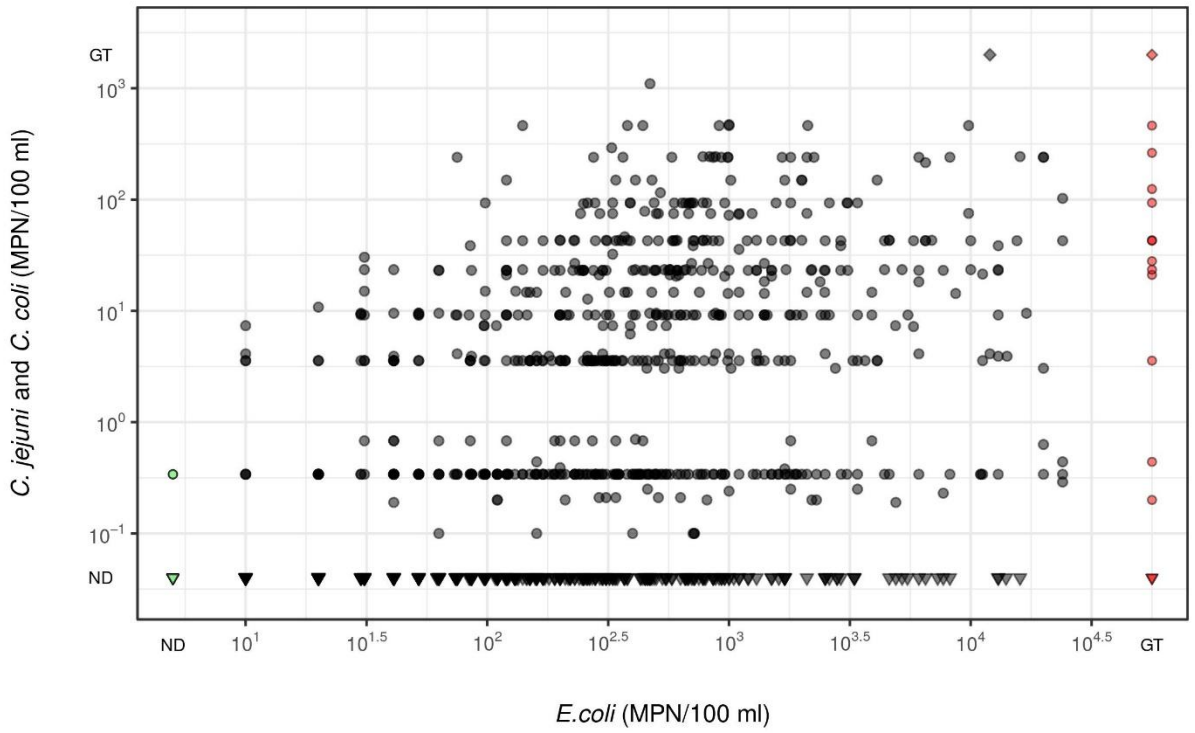
9.1 *CAMPYLOBACTER* AND *E. COLI*

This section will present the combined concentration data for *Campylobacter* that was identified as *C. jejuni* and/or *C. coli*. See section 7.2.3 for details of the other *Campylobacter* types identified in the study, most of which have yet to be associated with human illness.

9.1.1 All data

E. coli and *Campylobacter* concentration data was analysed from a total of 1,009 water samples, this excluded 20 water samples with a water salinity exceeding 1,000 µs/cm, and only included water samples where both *E. coli* and *Campylobacter* concentrations could be analysed (Figure 17). *E. coli* concentrations ranged from <10 to > 24,200 MPN/100 ml and combined *C. jejuni* and *C. coli* concentrations in the range < 0.1 to > 1,100 MPN/100 ml.

Figure 17: Comparison of *Campylobacter* (combined *C. jejuni and C. coli*) and *E. coli* concentrations in freshwater samples.



Notes:

E. coli concentrations below the detection limit of 10 MPN/100 ml (ND) are plotted in light green, and those above the limit of quantification of 24200 MPN/100 ml (GT) are plotted in red.

Campylobacter concentrations below the detection limit of 0.1 MPN/100 ml are plotted as triangles, and those above the limit of quantification of 1100 MPN/100 ml are plotted as diamond.

The banding of the *Campylobacter* and *E. coli* concentrations is due to the discrete nature of the MPN concentration estimates (McBride, 2005).

The plot does not indicate a predictive one to one relationship between *E. coli* and *Campylobacter* concentrations, where knowing the value of one of these bacteria will predict the value of the other for a single water sample within a close range of values. Instead, *Campylobacter* may be present or not across the range of *E. coli* concentrations observed, and within an interval of *E. coli* concentrations a range of *Campylobacter* concentrations can be seen.

The data in Figure 17 show that even at low *E. coli* concentrations, there can occasionally be elevated *Campylobacter* concentrations. For *E. coli* concentrations below 130 MPN/100 ml, a maximum of 240 MPN/100 ml of *Campylobacter* were observed in a single sample. For *E. coli* concentrations below 50 MPN/100 ml, *Campylobacter* concentrations stayed below 30 MPN/100 ml.

However, Figure 17 does indicate a possible trend of increasing *Campylobacter* concentrations with increasing *E. coli* concentrations. To explore trends in the data further the data was analysed by:

1. Grouping the water samples by *E. coli* concentration interval with roughly the same number of water samples in each group (Table 44) and describing the distribution of *Campylobacter* concentrations using box plots (Figure 18a). The boxplot boxes represent the central 25th to 75th percentile range of the *Campylobacter* concentrations, and the central line in the box is the median (50th percentile).

2. Applying quantile regression across the whole dataset to establish the 25th, 50th and 75th percentile regression estimates of the *Campylobacter* concentration distribution given the *E. coli* concentrations (Figure 18b). Quantile regression is a tool that can be used to show how the location and shape of the whole distribution of a variable changes conditional on how another variable changes (Koenker, 2005).

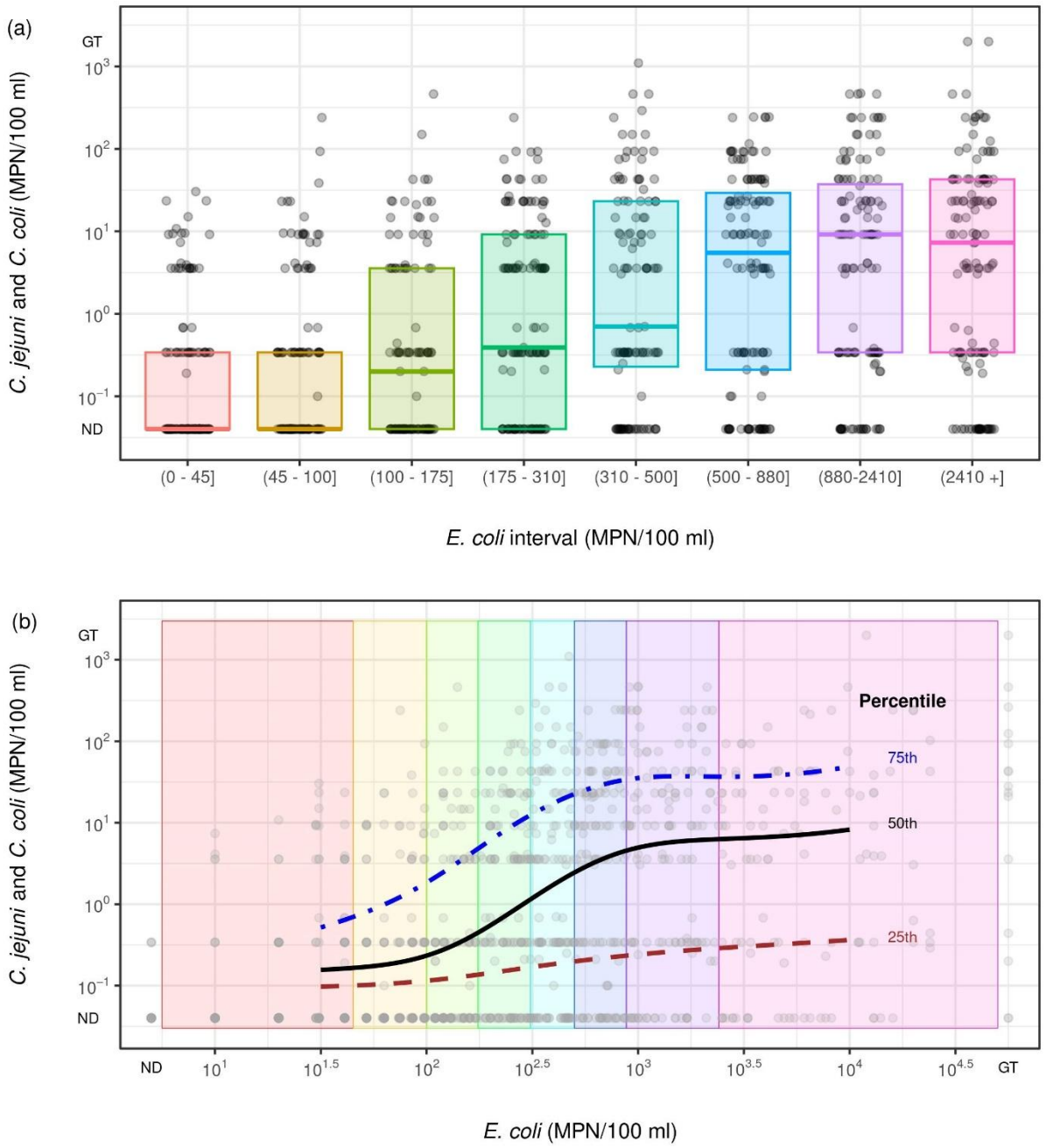
The data across the range of *E. coli* concentrations comes from a good proportion of the 71 sites that where sampled. The water samples associated with each of the *E. coli* intervals includes data from between 36 and 48 sites (Table 44).

Table 44: Number of samples, sites and detections of *C. jejuni* and/or *C. coli* in each *E. coli* concentration interval.

E. coli interval (MPN/100 ml)	Number of water samples	Number of sites	Percentage of samples with *Campylobacter* detected.
Up to 45	137	36	36%
46 to 100	125	43	46%
101 to 175	117	40	51%
176 to 310	137	48	71%
311 to 500	119	41	76%
501 to 880	124	40	79%
880 to 2410	124	39	83%
Above 2,410	124	43	80%

The visualisation of the data and interquartile range (25th to 75th percentile) of the *Campylobacter* distribution (Figure 18) shows the distribution location and shape does change with increasing *E. coli* concentration up to about 1000 *E. coli* MPN/100 ml. Above 1000 *E. coli* MPN/100 ml, there is little change to the *Campylobacter* concentration distribution.

Figure 18: *Campylobacter* concentration and interquartile range (25th to 75th percentile) by (a) *E. coli* concentration intervals with even sample size displayed as boxplots and (b) smoothed 25th, 50th and 75th percentiles using quantile regression.



Notes:

Boxplot box represents the interquartile range of values (25th to 75th percentiles) and horizontal bar is plotted at median (50th percentile) value. *Campylobacter* concentrations plotted as jittered individual concentrations. In panel (a) the values on the x-axis between (and] indicate a range that is greater than the first value, but less than or including the upper value.

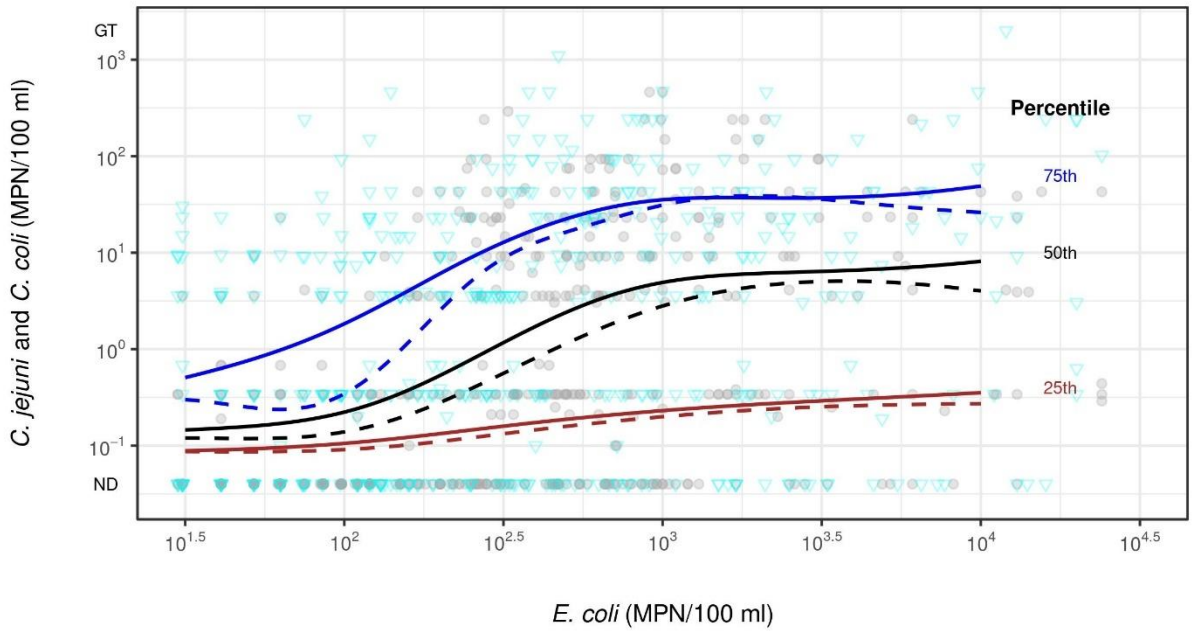
9.1.2 Bathing Season

The survey design was to collect data throughout the year. Of the 1,009 water samples with *Campylobacter* and *E. coli* concentration data, 491 (49%) were collected during a bathing season. For the purpose of this section the bathing season is defined by 31 October to 31 March[4], with the October 31 chosen rather than 1 November to include this sampling day.

The data collected in the bathing season had a different distributional profile compared to the full dataset (Figure 19). Fewer higher *Campylobacter* concentrations were observed at lower *E. coli* concentrations in the bathing season, than outside the bathing season. To see if this difference is significant in terms of the infection rates from swimming in the water. The simulation model (section 10 to 13) will consider models based on water samples collected all year round and models based on water samples collected from 31 October to 31 March.

Multivariate analysis of environmental factors suggests that *Campylobacter* concentrations decrease with increasing water temperature (Appendix E) which may account for the difference in the distributions of *Campylobacter* concentrations seen in the bathing season compared to cooler months.

Figure 19: Comparison of *Campylobacter (C. jejuni and C. coli)* and *E. coli* concentrations in freshwater samples during (circles) and outside (triangles) the bathing season. Solid lines are the quantile regression estimates using all the data and the dashed lines are the quantile estimates based on data collected during the bathing season (31 October to 31 March).



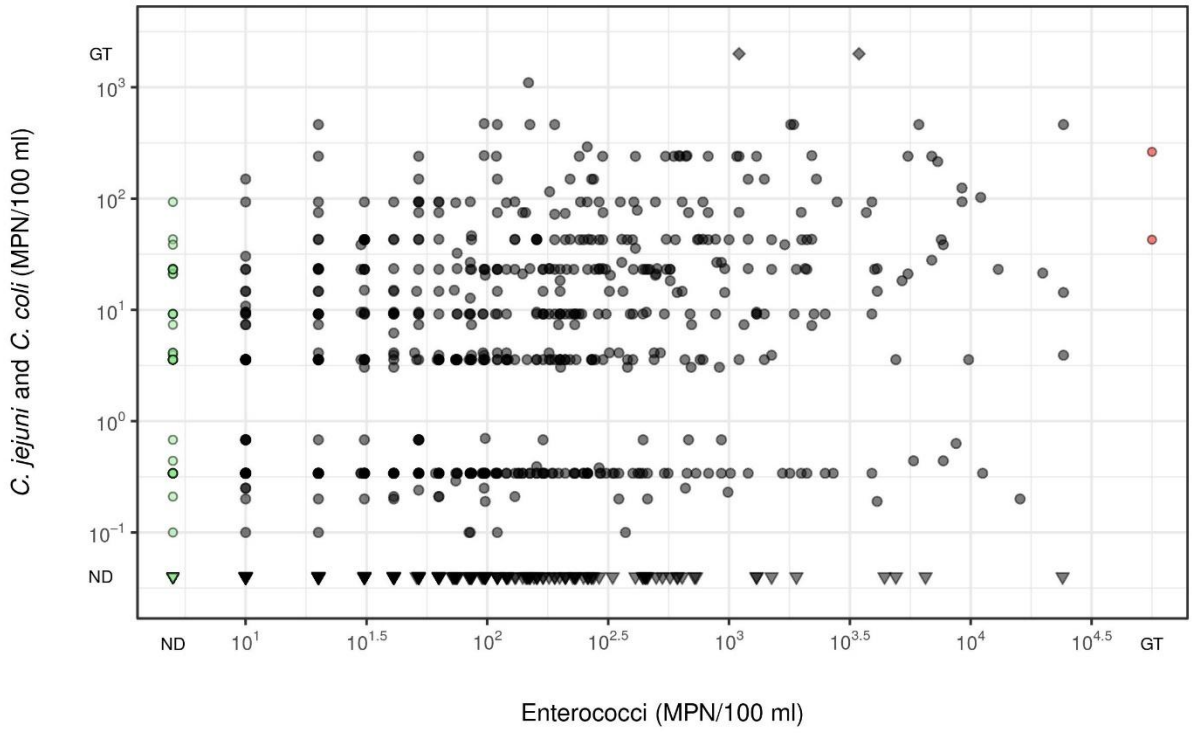
[4] The bathing season for each primary contact site is defined by the regional council based on the time period during the year that a site would be regularly used. So will vary throughout the country (*National Policy Statement for Freshwater Management 2020*. - updated October 2024).

9.2 *CAMPYLOBACTER* AND ENTEROCOCCI

This section will present the data relationship between the combined concentration data for *Campylobacter* that was identified as *C. jejuni* and/or *C. coli* and the faecal indicator bacteria, enterococci.

Enterococci and *Campylobacter* concentration data was analysed from a total of 1,028 water samples (Figure 20). Enterococci concentrations ranged from <10 MPN/100ml (153 water samples) to > 24,200 MPN/100 mL (two water samples) and combined *C. jejuni* and *C. coli* concentrations in the range < 0.1 to > 1,100 MPN/100 mL. In comparison with *E. coli* concentrations, enterococci had a lower prevalence of samples with concentrations above the detection limit and generally lower concentrations than *E. coli* (section 4.1).

Figure 20: Comparison of *Campylobacter (C. jejuni and C. coli)* and enterococci concentrations in freshwater samples.



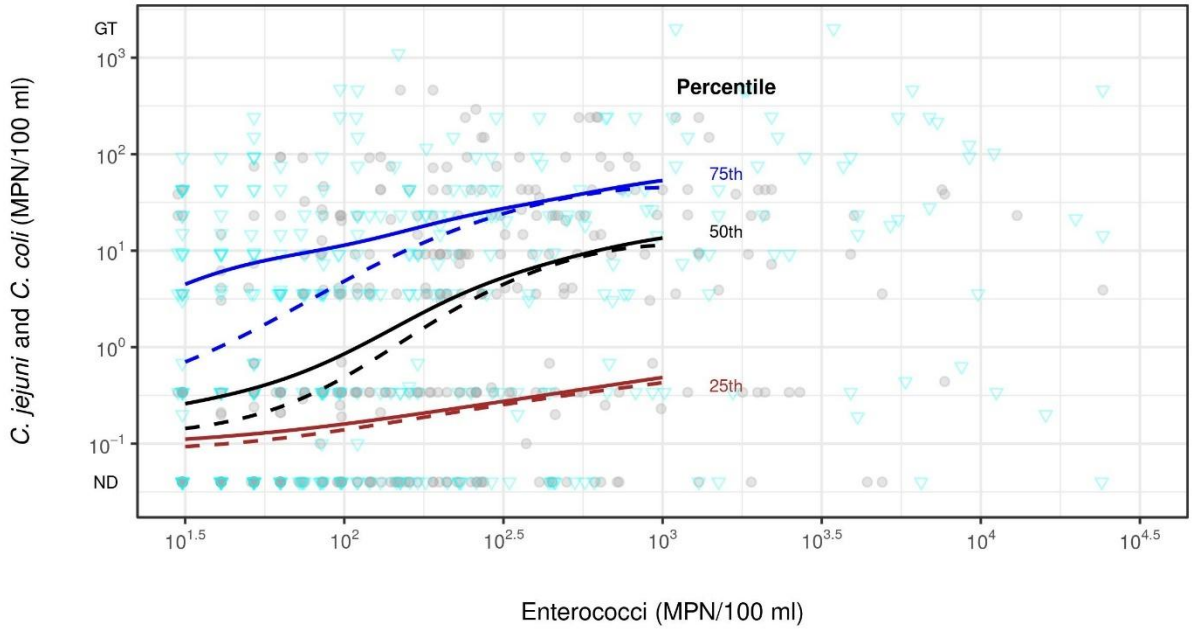
Note: Enterococci concentrations below the detection limit of 10 MPN/100 ml (ND) are plotted in light green, and those above the limit of quantification of 24,200 MPN/100 ml (GT) are plotted in red. *Campylobacter* concentrations below the detection limit of 0.1 MPN/100 ml are plotted as triangles, and those above the limit of quantification of 1,100 MPN/100 ml are plotted as a diamond.

Similar to the relationship between *E. coli and Campylobacter,* Figure 20 does not indicate a predictive one to one relationship between enterococci and *Campylobacter* concentrations, where knowing the value of one of these bacteria will predict the value of the other for a single water sample within a close range of values. Instead, *Campylobacter* may be present or not across the range of enterococci concentrations observed, and within an interval of enterococci concentrations a range of *Campylobacter* concentrations can be seen.

The data in Figure 20 show that even at low enterococci concentrations, there can occasionally be *Campylobacter* present at a concentration that could cause illness in humans. For enterococci concentrations of 10 MPN/100 ml or less (shown as 10^1 in Figure 20), the *Campylobacter* concentration ranged from < 0.1 to 150 MPN/100 ml.

However, Figure 20 does indicate a possible trend of increasing *Campylobacter* concentrations with increasing enterococci concentrations. To explore trends in the data further the data was analysed by applying quantile regression across the whole dataset or just data from the bathing season (31 October to 31 March). The 25th, 50th and 75th percentile regression estimates of the *Campylobacter* concentration distribution given the enterococci concentrations up to 1,000 MPN/100 ml[5] are shown in Figure 21.

Figure 21: Comparison of *Campylobacter (C. jejuni and C. coli)* concentrations with enterococci concentrations in freshwater samples during (circles) and outside (triangles) the bathing season. Solid lines are the quantile regression estimates using all the data and the dashed lines are the quantile estimates based on data collected during the bathing season (31 October to 31 March).



Similar to the relationship between *E. coli* and *Campylobacter* concentrations in a water sample, there was an increasing trend in *Campylobacter* concentrations with increasing enterococci concentration. The largest differences in the percentiles of the *Campylobacter* concentrations between the whole dataset and only data from the bathing season, was again at the lowest enterococci concentrations.

[5] There were too few water samples with enterococci above 1,000 MPN/100 ml to robustly use in the quantile regression. Thirty-eight water samples were taken during the bathing season, and 51 water samples outside the bathing season.

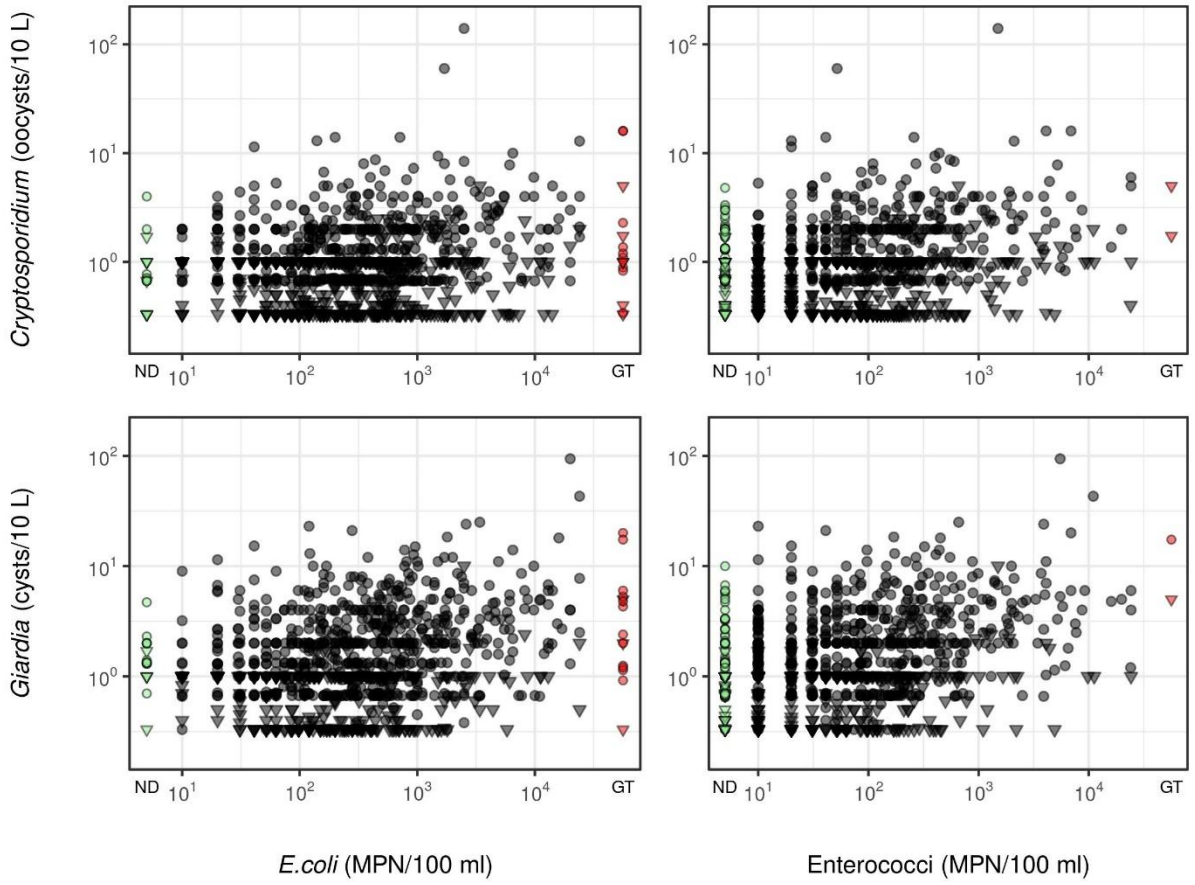
9.3 CRYPTOSPORIDIUM AND GIARDIA

9.3.1 All data

Data with *Cryptosporidium* or *Giardia* concentration with *E. coli* concentration was available for 989 water samples and with enterococci for 983 water samples. This includes all data where 1 litre or more of water could be analysed. The maximum concentrations for protozoa were 140 *Cryptosporidium* oocysts/10 litre and 94 *Giardia* cysts/10 litre.

Similar to conclusions observed for *Campylobacter*, there is no clear predictive one to one trend of *Cryptosporidium* or *Giardia* concentration with *E. coli* or enterococci concentration (Figure 22).

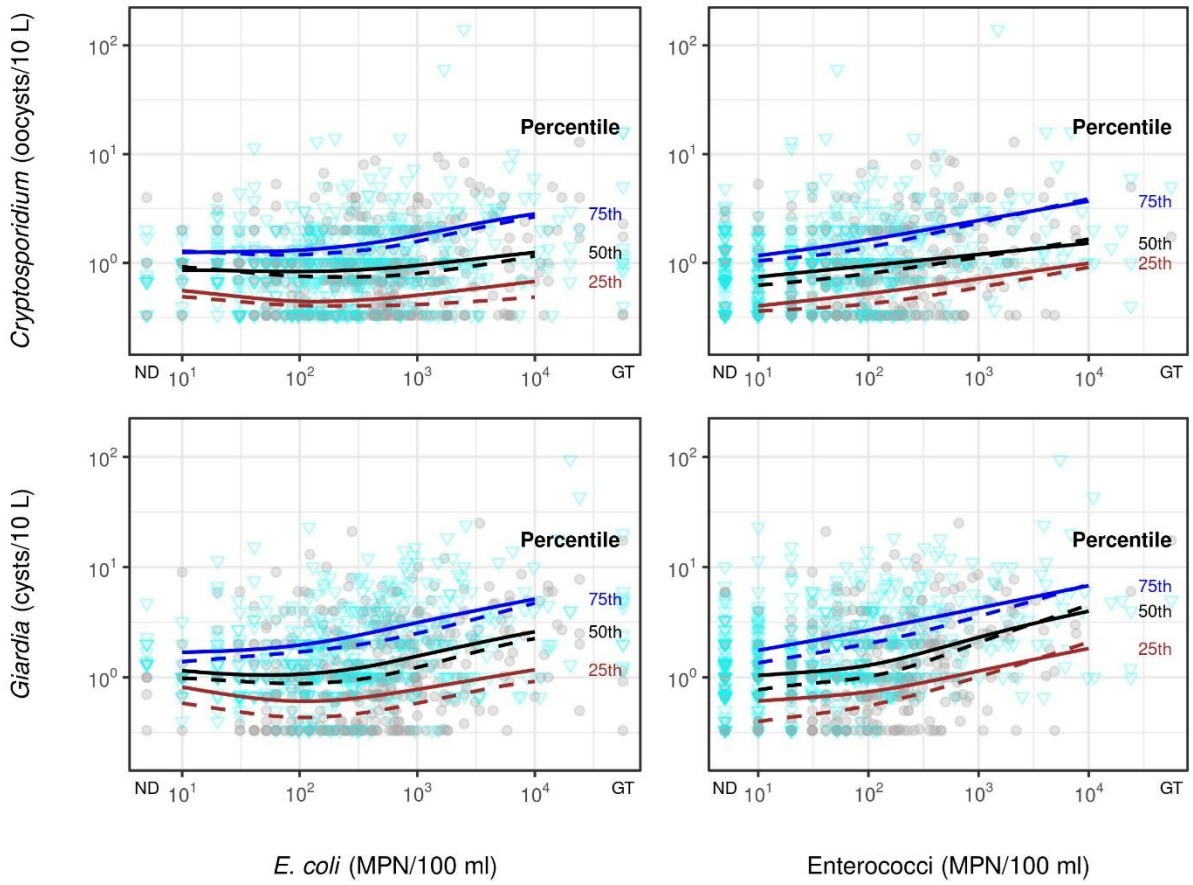
Figure 22: Comparison of *Cryptosporidium* and *Giardia* concentrations with *E. coli* or enterococci concentrations in freshwater samples.



Note: *E. coli* and Enterococci concentrations below the detection limit of 10 MPN/100 ml (ND) are plotted in light green, and those above the limit of quantification of 24,200 MPN/100 ml (GT) are plotted in red. *Protozoa* concentrations below the detection limit are plotted as triangles, the detection limit is dependent on the volume of water that was filtered.

Applying quantile regression to the data, shows *E. coli* or enterococci are only a weak predictor of increasing *Cryptosporidium* or *Giardia* concentrations (Figure 23). There is also minimal difference between the data compared to using data collected in the bathing season, as summarised by 25th, 50th and 75th percentiles.

Figure 23: Comparison of *Cryptosporidium and Giardia concentrations compared to E. coli or* enterococci concentrations in freshwater samples during (circles) and outside (triangles) the bathing season. Solid lines are the quantile regression estimates using all the data and the dashed lines are the quantile estimates based on data collected during the bathing season (31 October to 31 March).



9.4 FIELD VARIABLES AND MST AS PREDICTORS OF CAMPYLOBACTER CONCENTRATION

This section provides an initial analysis to investigate the ability of field variables and MST markers to explain some of the variability of in *Campylobacter* concentration values at different *E. coli* and enterococci concentrations, based on the nationally collected data.

9.4.1 Modelling Approach

Linear mixed effects models ("gamm" models from the R mgcv package (Wood 2017)), with multiple independent variables was analysed using $\log_{10}$ *E. coli* or enterococci concentrations, $\log_{10}$ MST marker concentrations and field variables as the independent variables. To ensure the range of values of the different variables did not bias the modelling, the model was run with the independent variables scaled and centred by median values before applying the regression analysis.

As in section 8, the FIB, MST and field variables were modelled as fixed independent variables X_i and the between site variability, $f(site)$, measured as a random effect in the form of a normal distribution (mean zero and standard deviation derived from the model fitting). The *Campylobacter* concentration at observation j is modelled by:

$$\log_{10} Campylobacter\ concentration_j = \sum(\beta_i \times X_{ij}) + f(Site) + \varepsilon . \qquad [9.1]$$

The full model including all the possible variables was first examined to determine which variables were predictive of *Campylobacter* contribution. This was followed by a stepwise deletion of variables based on model fit (Akaike information criterion (AIC) and log likelihood) until no further improvement was observed to create a base model.

Two alternative model formats were then compared against the base model. Firstly, potential interaction terms were included and secondly, additive models which do not require the independent variable to vary linearly with $\log_{10}$ *Campylobacter* concentration. Additive models replace the $\beta_i X_i$ component of the model with a smoothing function $s(X_i)$. For each independent variable in the base model, an additive model was run with a smoothing function applied to the independent variable.

The models were run using either *E. coli* or enterococci as the FIB independent variable.

9.4.2 Results

The best fitting models from the above process, were the simplest models, with no additive smoothing components or interaction terms. When additive models were trialled, the best fitting smoothing functions were approximately linear, confirming the appropriateness of the linear model.

The best fitting base model consists of:

- *E. coli* or enterococci, ruminant (BacR) and avian (GDF) MST marker $\log_{10}$ concentrations which all contribute to increasing *Campylobacter* concentration estimates with increases in their concentrations.
- Water temperature and dissolved oxygen which both contribute to decreasing *Campylobacter* concentration estimates with increases in their values.

For the national level dataset, the other independent variables; human MST markers, rainfall, turbidity, specific conductivity and pH were not significantly predictive of *Campylobacter* concentration. These variables should not be ruled out from modelling specific sites. The combined random site effect (standard deviation ~0.21) and residual (standard deviation ~0.56) suggest there are factors not included in the base model which are associated with changes in *Campylobacter* concentrations.

Water temperature is weakly correlated to dissolved oxygen in the modelling dataset ($R^2 = 0.4$), but the inclusion of both improved the model predictions. Recall that water temperature is seasonal (appendix D, Figure 40) and the connection with *Campylobacter* concentration may be at least partially due to other seasonal factors such as UV intensity and duration at higher intensity, which may affect die off rates in clear fresh water. *Campylobacter* has been shown to survive better at the lower water temperature observed in this study, than at higher temperatures (Buswell et al, 1998).

This freshwater study has generated a large set of data which can be further analysed for relationships for different subgroups of the data. Furthermore, it provides support information for the setting of local or national targets or when intervention strategies are under consideration.

PART C

Quantitative Microbial Risk Assessment

(Campylobacter, Cryptosporidium and Giardia)

10 INTRODUCTION TO THE QUANTITATIVE MICROBIAL RISK ASSESSMENT SIMULATION MODEL

10.1 QUANTITATIVE MICROBIAL RISK ASSESSMENT FRAMEWORK

Quantitative microbial risk assessment provides a systematic and transparent approach to providing an evidence-based health risk assessment. Quantitative microbial risk assessments use mathematical models that use quantitative data to describe the scenario being assessed and produce quantitative health risk estimates.

Quantitative microbial risk assessment models can be developed at various levels of detail or complexity. They can range from high level models, which can be used to provide evidence for supporting national guideline setting, to more detailed freshwater site models that councils could use to compare the relative merits of different strategies to make freshwater sites safer for recreational use.

This project followed the four step QMRA framework as suggested in the WHO quantitative microbial risk assessment: application for water safety management guidance document (World Health Organization, 2016) as outlined in Figure 24.

10.2 PROBLEM FORMULATION

The 2003 Guidelines were developed to help water managers monitor, report on and control the public health risk posed by microbiological contamination of recreational waters.

The freshwater component of the Guidelines uses the indicator organism *E. coli* to assess water quality, with numeric guideline values developed from the findings of the QMRA from the 1998-2000 FMRP (McBride et al, 2002). There have been significant changes over the last 20 years in land-use practices, laboratory methods, computation power and data on human behaviours during recreational water activities. New tools such as MST have been developed which provide a greater understanding of the sources of the faecal contamination found in rivers than was available during the FMRP.

The purpose of the current QMRA simulation model is to enable use of the new survey data collected between 2020 and 2024 across New Zealand to provide human health risk estimates from recreational water activities and explore how risks are associated with the concentration profiles of indicators. Indicators include *E. coli*, enterococci, and MST markers.

The QMRA used Monte Carlo simulation models (Vose, 2008) which allow the replacement of point values of model components with statistical distributions representing observed variability. Uncertainty in the risk estimates due to uncertainty in the dose-response relationships was also analysed by two-dimensional Monte Carlo simulations. The effect of uncertainty in the other model parameters on the model outputs was considered via sensitivity analysis.

The scope of the QMRA was defined by:

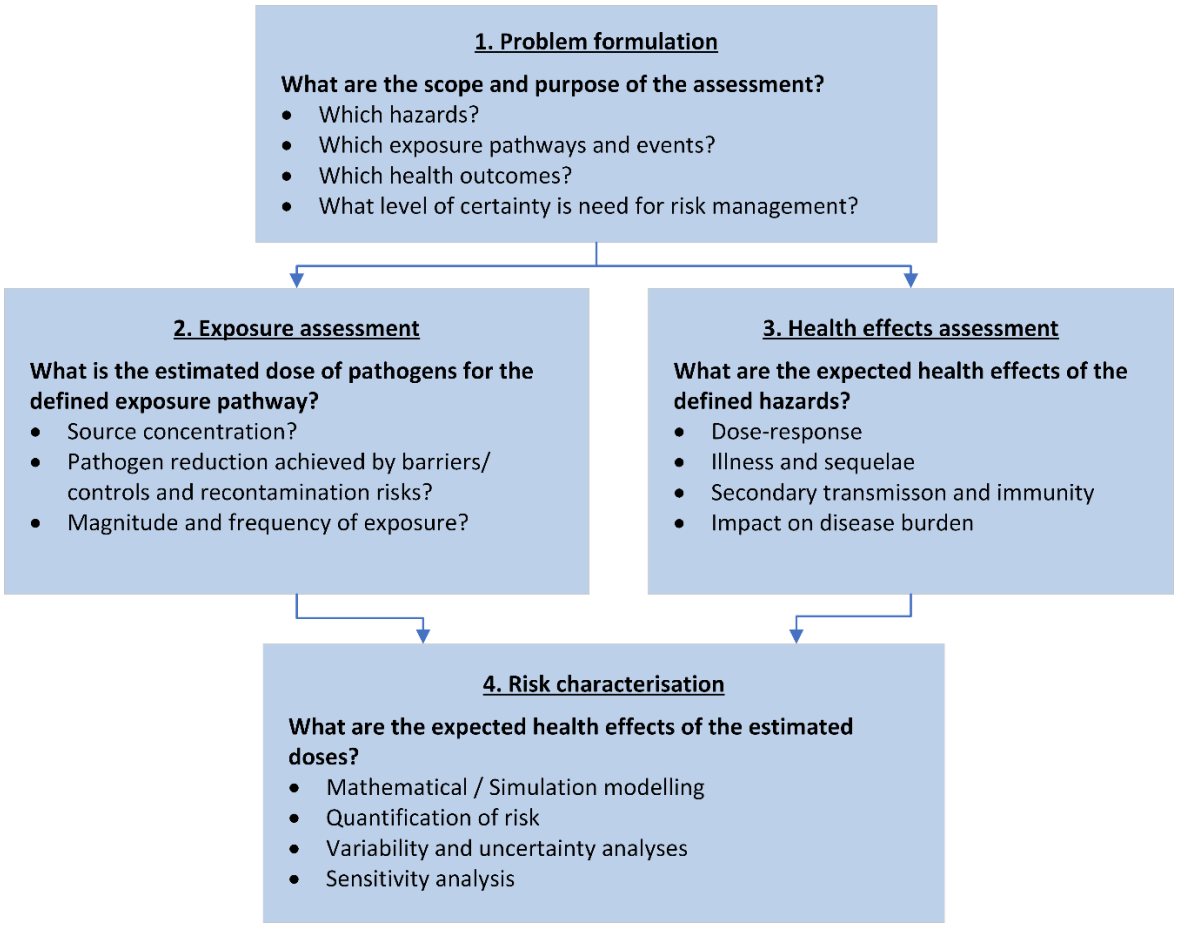
Hazards included: The water sampling data provided concentration data for three pathogenic microorganisms: *Campylobacter (C. jejuni and C. coli)*, *Cryptosporidium* and *Giardia*. These three microorganisms are known to be present at freshwater sites, the resultant infections are notifiable diseases in New Zealand, quantitative measures of concentrations in freshwater in the 2020 – 2024 study data are available, and published dose-response relationships are available that could be used in the QMRA.

Exposure pathways: Swallowing water while swimming at a freshwater site. The water may be contaminated with pathogens from human, animal or bird faecal sources.

Health outcomes: For each pathogen and *E. coli* or enterococci concentration in the water;

- The population risk of acute gastroenteritis described by the percentage of swimming events which result in infection or illness, and the DALYs due to 1000 swims.
- The distribution of risk to individual swimmers.

Figure 24: Quantitative microbial risk assessment framework. Adapted from World Health Organization (2016).



11 EXPOSURE ASSESSMENT

11.1 EXPOSURE PATHWAY

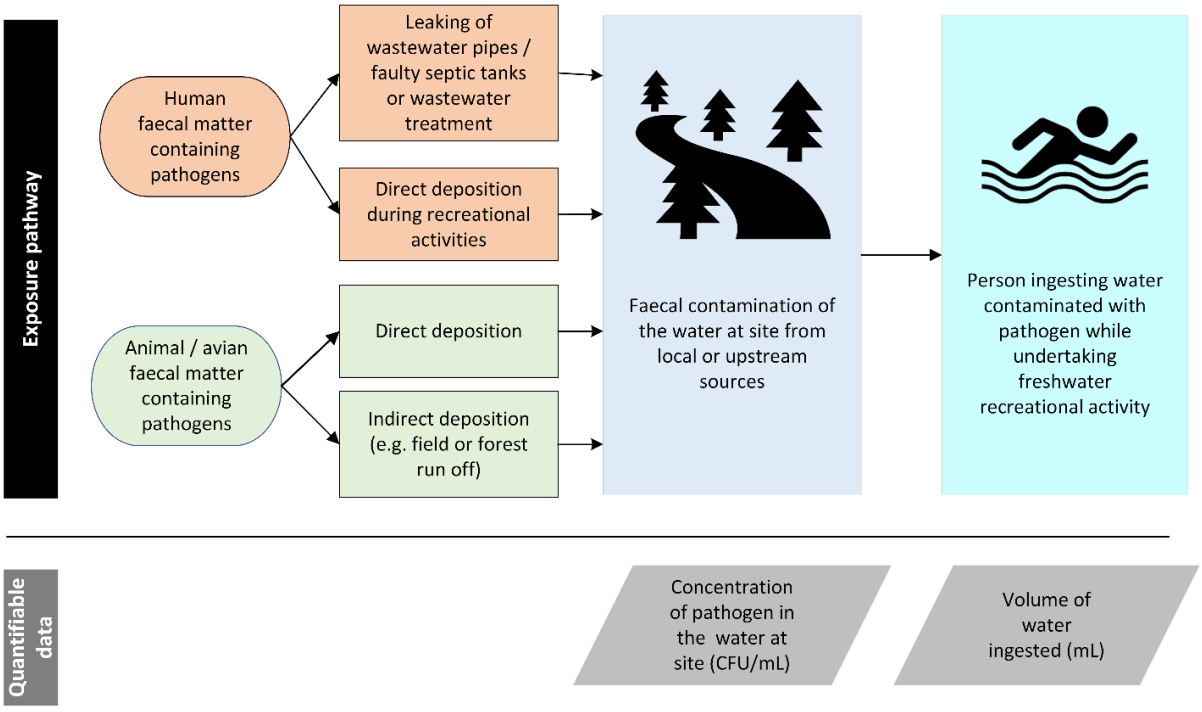
The pathway for exposure of people to pathogens from freshwater activities is summarised in Figure 25.

Faecal matter can enter waterways by direct deposition into the water, by transfer of matter into waterways from land runoff during rainfall or from wastewater infrastructure. Faecal matter will disperse in the river water, with dilution reducing the concentration of faecal matter downstream from the contamination point.

Pathogens such as *Campylobacter*, *Salmonella*, STEC, *Giardia* and *Cryptosporidium* can be found in animal and human faecal sources. Most viruses are species-specific so human infective viruses such as norovirus will only be found in human faecal sources.

People undertaking recreational water activities may ingest water. If the water contains pathogenic microorganisms, they will also ingest a dose of these microorganisms, which may be sufficient to establish an infection and potentially cause illness.

Figure 25: Exposure pathway for single freshwater swimming event



Swimmer and river graphics from https://www.vecteezy.com/

11.1.1 Assumption

The quantification of pathogens occurs at the site where swimming and water ingestion occurs and represents infective organisms. The sample microorganism concentrations at a site are representative of the concentrations in the water ingested by a recreational water user.

Knowledge of the initial faecal contamination concentrations or the dilution that occurs between the source of the faecal contamination and the swimming site is not required.

11.2 PATHOGENS AT FRESHWATER SITES

Parts A and B of this report have shown that *C. jejuni and coli*, *Cryptosporidium* spp. and *Giardia* spp. were detected in the 2020 to 2024 study water samples and therefore have the potential to cause illness in swimmers at the sampling sites where they are present. *C. jejuni* and *C. coli* were detected in 64% and 13% of the water samples respectively at a detection limit of 0.1 MPN/100 ml (Table 20).

Cryptosporidium was present in 50% of samples and *Giardia* in 69% of samples (Table 31).

11.3 ESTABLISHING THE RELATIONSHIP BETWEEN *CAMPYLOBACTER* AND FIB CONCENTRATIONS

The generation of *Campylobacter* concentrations for a given *E. coli* or enterococci concentration in the simulation model is defined in a three-step process:

1) *Campylobacter* concentration values are estimated which account for MPN measurement uncertainty.

2) The percentiles of the simulated *Campylobacter* concentrations given FIB concentration are determined by quantile regression.

3) For each target FIB concentration, the *Campylobacter* concentrations are sampled from the distribution defined by the percentiles derived in step 2.

11.3.1 Step one – MPN uncertainty

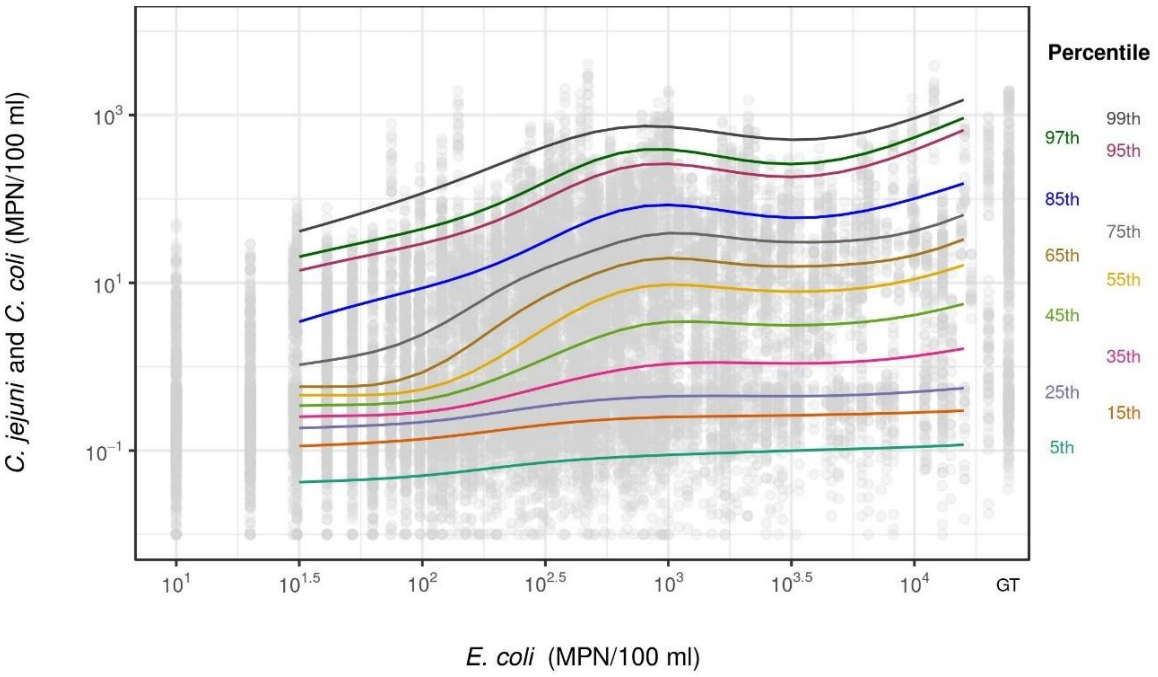
The MPN enumeration approach leads to discrete *Campylobacter* concentrations being recorded for water samples, as evident in Figure 17 and Figure 18 (McBride et al, 2002). The MPN values provide the most probable concentration given observed growth of *Campylobacter* in the dilution tubes (Appendix B.1.2), but there is associated measurement uncertainty around the reported MPN values.

The software used to calculate the MPN values, also provides an estimate of the natural log normal distribution of the measurement uncertainty. To reflect this measurement uncertainty, for each water sample, 20 *Campylobacter* concentrations were sampled from the uncertainty distribution for each MPN value, resulting in 20 x 1,009 data points for describing the relationship between *Campylobacter* and *E. coli* concentrations (Figure 26). Using more than 20 uncertainty values per water sample did not improve the model estimates.

11.3.2 Step two – Estimating the *Campylobacter* distribution for *E. coli* concentrations

To explore the distribution of concentrations, quantile regression was applied as described in Appendix B.5.2, at the 5th to 95th percentiles of simulated *Campylobacter* concentrations in 10 percent intervals (Figure 26). The highest percentiles are at the highest $\log_{10}$ *Campylobacter* concentrations where small changes in the plot relate to larger changes in concentrations due to the log scale. Examining the percentiles between 95th and 99th percentiles (97th and 99th percentiles are shown in Figure 26) suggests the distribution above the 95th percentile should be represented by a triangular distribution to provide an upper tail.

Figure 26: Quantile regression estimates for *Campylobacter* distributions with different *E. coli* concentrations



Comparison of the estimated 5th to 99th percentile values with the simulated data showed there was good agreement. For all but one percentile line, there was less than 0.5% percent difference between the number of observations below the fitted percentile line and the target percentage. The largest difference was for the 65th percentile, where 66% of observations were below the fitted percentile line.

As observed in the original data, *Campylobacter* concentrations below the 5th percentile all related to results that were less than the limit of detection.

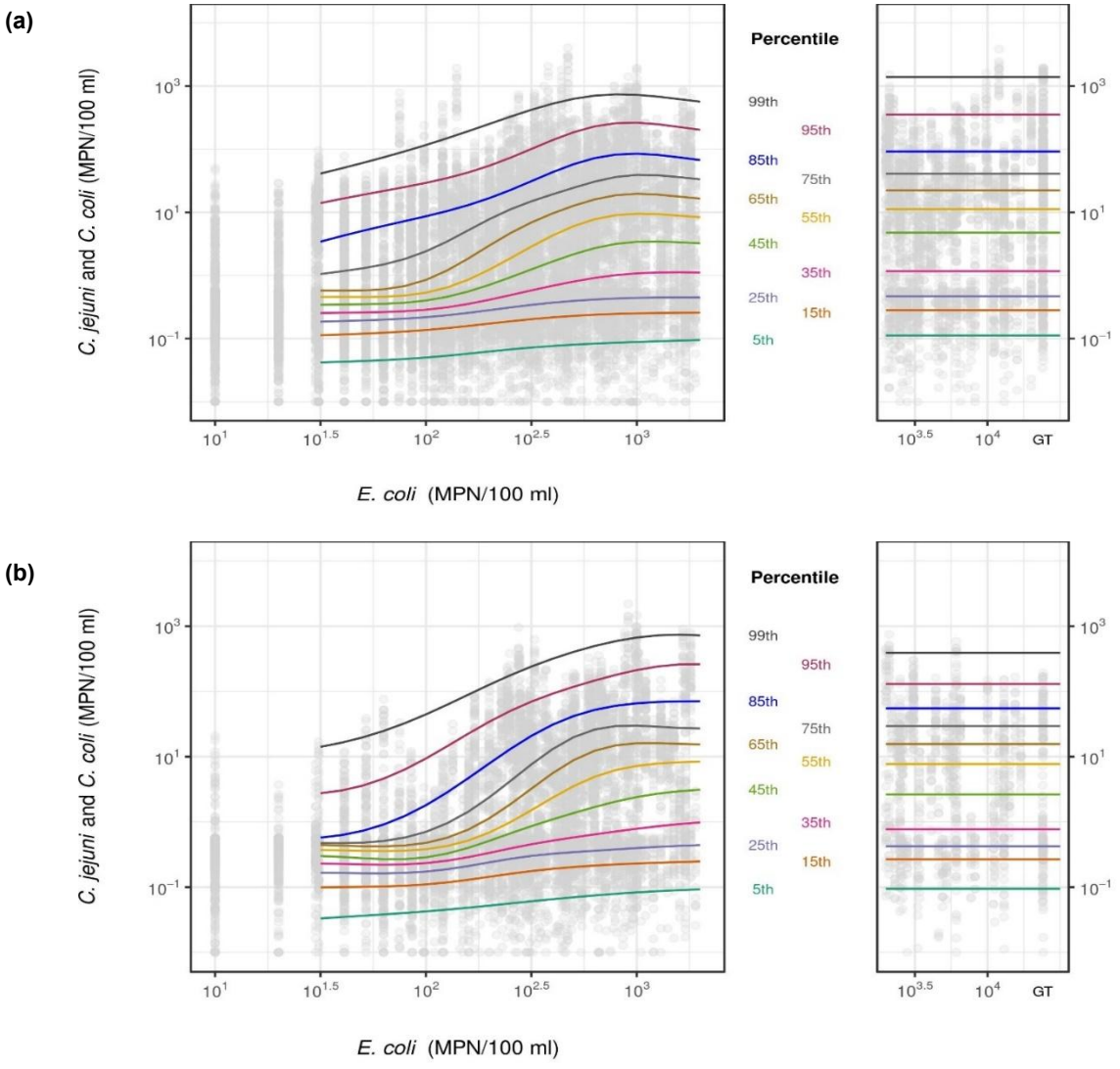
The percentile lines provide a suitable framework to model how the distribution of *Campylobacter* concentrations changes with *E. coli* concentration. However, the percentile lines indicate a slight dip in the 50th to 99th percentiles of *Campylobacter* concentrations for water samples with *E. coli* concentrations in the range 2500 to 10,000 MPN/100 ml ($10^{3.4}$ to 10^4). There is no obvious microbiological or hydrological reason for this non-monotonic slope, and it is likely the lower than expected *Campylobacter* concentrations in these water samples are due to fewer water samples with *E. coli* concentrations in this range.

Consequently, the simulation model of the *Campylobacter* distribution was split into two components (Figure 27):

i. *Campylobacter* concentration distributions based on the quantile regression of the water samples with *E. coli* concentrations up to $10^{3.4}$ MPN/100 ml.

ii. The remaining set of water samples with *E. coli* concentrations above $10^{3.4}$ MPN/100 ml was modelled as a single group, resulting in a single *Campylobacter* concentration distribution.

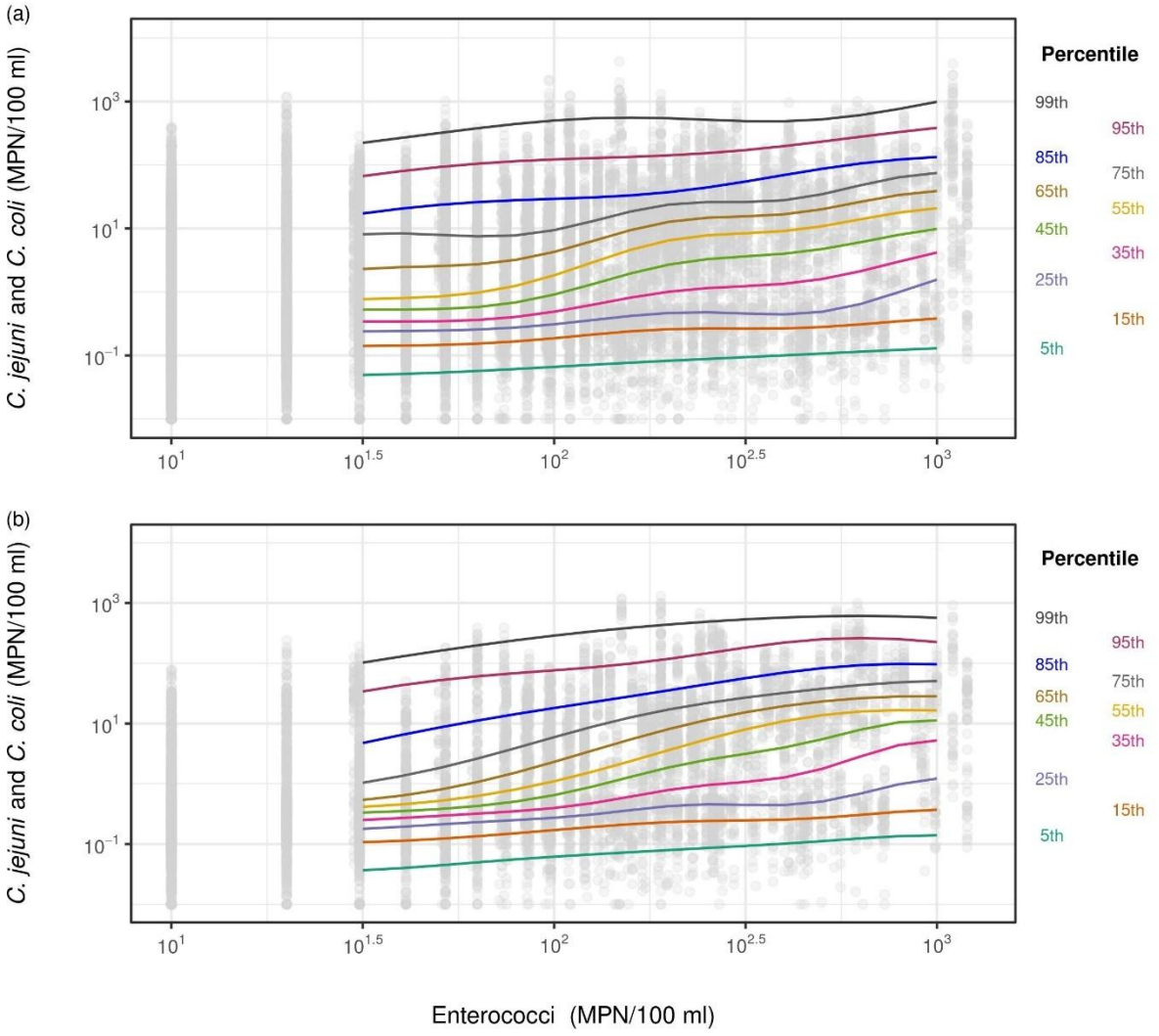
Figure 27: Quantile regression estimates for *Campylobacter* distributions with *E. coli* concentrations using (a) all data and (b) bathing season data only



11.3.3 Step two – Estimating the *Campylobacter* distribution for enterococci concentrations

Quantile regression was applied to the *Campylobacter* distributions for water samples with enterococci concentrations below 1000 MPN/100 ml (section 9.2). Quantile regression was applied as described in Appendix B.5.2, at the 5th to 95th percentiles of simulated *Campylobacter* concentrations, in 10 percent intervals, and at the 99th percentile as shown in Figure 28.

Figure 28: Quantile regression estimates for *Campylobacter* distributions with enterococci concentrations using (a) all data and (b) bathing season data.



11.3.4 Step three – Sampling from the *Campylobacter* distribution

The *Campylobacter* concentrations for the simulation model are calculated in the following way for each target FIB (*E. coli* or enterococci) concentration:

- Five percent of samples are set to zero *Campylobacter* concentration.

- Ten percent of samples are set in each of the 5-15, 15-25, … 85-95 regression percentile intervals for *Campylobacter* at the given FIB concentration. For each interval, a uniform distribution is used to sample between the interval bounds.
- Five percent of samples are sampled from a triangle distribution between the 95th to 99th percentile. With the most likely value at the 95th percentile.

11.4 ESTABLISHING THE RELATIONSHIP BETWEEN PROTOZOA AND ENTEROCOCCI

The simulation model results for *E. coli* are not presented in this report, given the similarity in the relationships between *E. coli* and enterococci and the two protozoa.

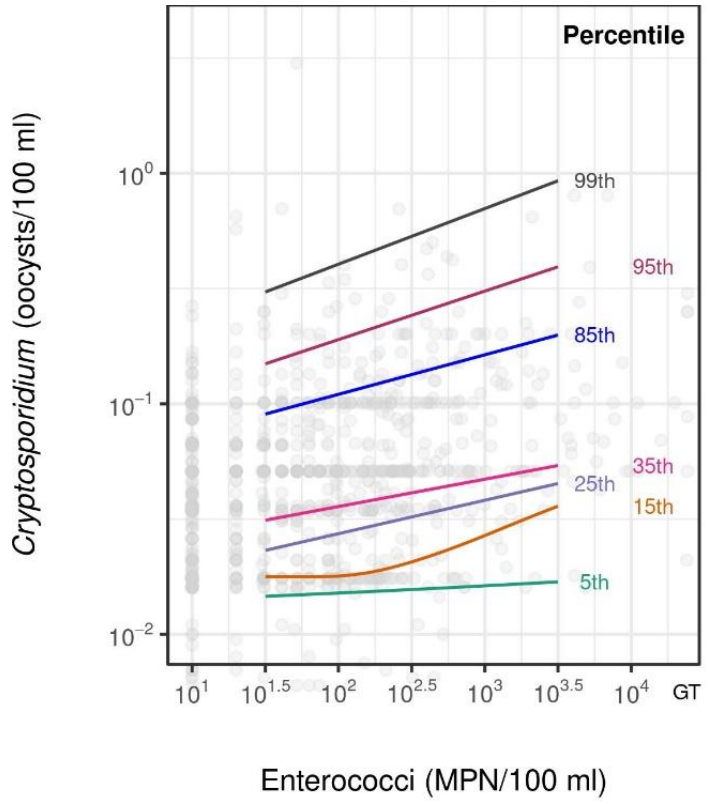
The relationship between *Cryptosporidium or Giardia* with enterococci in the simulation model was defined in a three-step process:

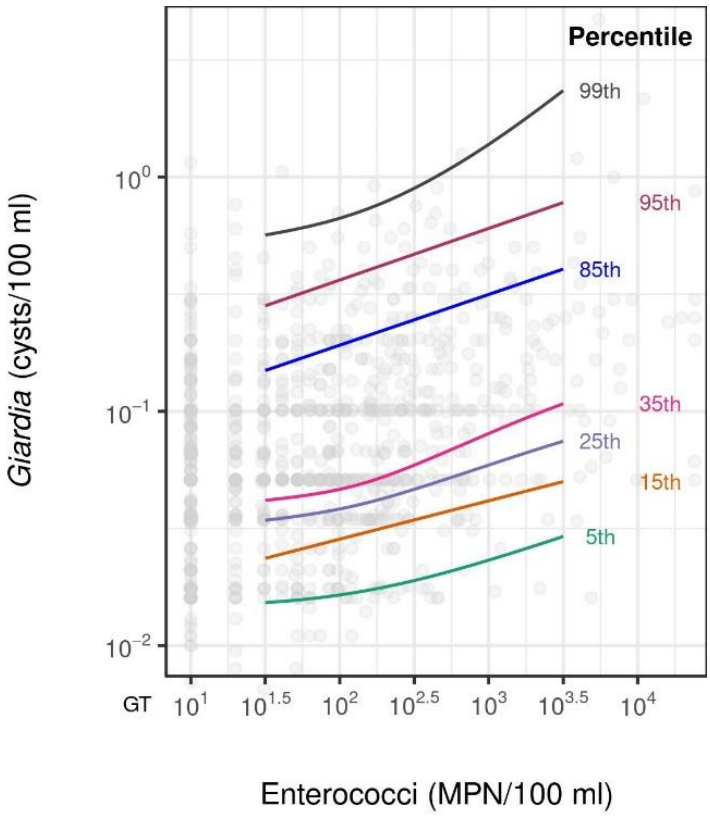
1. *Cryptosporidium or Giardia* concentration values were adjusted to account for laboratory measurement sensitivity (recovery rates). Concentrations were adjusted for a worst case scenario of 20% recovery rate (section 7.5.1) and the concentrations of samples with not detected results were assigned a value equal to the limit of detection of the method, based on the volume of water filtered.
2. Quantile regression was applied to the simulated *Cryptosporidium or Giardia* concentrations.
3. For each enterococci concentration, the *Cryptosporidium* or *Giardia* concentrations are sampled from the distribution defined by the percentiles derived in step 2

Quantile regression for the 5th to 95th and 99th percentiles are shown in Figure 29. To translate the regression lines at a given enterococci concentration to a distribution of *Cryptosporidium* or *Giardia* the following approach was taken:

- Five percent of samples were sampled from a uniform distribution with minimum value of 0.001/100mL and maximum value of the 5th percentile.
- Ten percent of samples were sampled from uniform distributions between 5-15th, 15-25th, 25-35th, 85-95th percentile intervals and 50% of samples from the 35-85th percentile interval (total of 90 percent of samples).
- Five percent of samples were sampled from a triangular distribution, between the 95th and 99th percentile. To allow for rare higher concentrations in this 5 percent of samples, 0.4% and 0.2% of samples for *Cryptosporidium* and *Giardia* were sampled from between 7 and 10/100 ml.

Figure 29: Quantile regression estimates for *Cryptosporidium* and *Giardia* distributions with different enterococci concentrations





11.5 RECREATIONAL ACTIVITIES CONSIDERED

Cressey and Horn compared published estimates of water ingestion rates and duration of exposure for different recreational water activities (Cressey and Horn, 2016). The greatest volumes of water were ingested during leisure swimming, scuba diving and surfing activities compared to other activities such as wading, boating or fishing. The estimated ingestion of freshwater by scuba divers in the Netherlands (Schijven and de Roda Husman, 2006) was similar to that of child swimmers in the United States (Dufour et al, 2017). In New Zealand, surfing and scuba diving are predominantly carried out in marine environments. As such, leisure swimming was considered to be the freshwater recreational activity resulting in the greatest potential pathogen exposure.

A New Zealand based freshwater survey conducted in the summer of 1996 interviewed 1,642 people visiting one of 11 North Island freshwater recreational sites, including 430 people under 13 years of age. Among those surveyed, 68% entered the water, with 54% swimming and 14% just paddling. Of those who swam, 95% indicated they put their head under the water (McBride et al, 1996). These findings suggest that full immersion swimming was a common activity at freshwater sites at this time.

A survey of year 11 students (mean age 15.6 years) from 41 high schools across New Zealand in 2000 found 64% had swum in a lake, pond or waterhole and 47% had swum in a river or creek in the previous year (Moran, 2009). A Department of Conservation 2020 survey of adults (aged 18 years and above) from all over New Zealand, found 28% of the 3000 respondents had been swimming in a lake or river in the last year (Ipsos Public Affairs, 2020). Fresh water swimming was recorded by more respondents than other freshwater activities. However, this survey may have been impacted by Covid-19 restrictions in 2020.

These studies confirm that freshwater swimming is an appropriate exposure scenario for a New Zealand focused QMRA.

11.6 POPULATION GROUP CONSIDERED

According to the 2020 Active New Zealand survey[6] conducted by Sport NZ, primary school children are more likely to go swimming than other age groups.

A pooled analysis of four summer epidemiological studies at freshwater temperate beaches in the USA found children in the 4 – 12 age group generally spent more time in the water than other age groups (DeFlorio-Barker et al, 2018). Of the people who contacted water during their beach visit, the 4 – 12 age group was also more likely to ingest water (53%) compared to 35% of the 13 – 18 age group and 16% of the 19 and older age group.

An USA study of swimmers in outdoor pools, found higher rates of water ingestion in 6 – 15-year-olds compared to adults aged 16 years and over (Dufour et al, 2017). A Dutch study asking people to recall their freshwater activities over the past year also suggested those aged under 15 are likely to ingest more water than the 15+ age group. However, this study is likely to be subject to recall bias which could be different for different ages (Schets et al, 2011).

A retrospective study of illness in 80,000 beach goers from 13 United States beach sites, children 12 years old and younger became ill at a greater rate with increasing concentrations of enterococcus spp. (odd ratio than adults aged 18 years or older (Wade et al, 2022). This effect was greatest for sites potentially impacted by human faecal contamination. The study concluded “children were at a higher risk of illness associated with faecal contamination as measured by indicator bacteria Enterococcus”. They also observed a relationship between the intensity of the swimming exposure and swimming-associated illness.

No studies were found which contradicted the above findings. Therefore, the QMRA inputs for the water ingestion rate and duration of activity was aligned to children.

11.7 HOW MUCH WATER IS LIKELY TO BE INGESTED?

The volume of water ingested during freshwater recreational activities is based on two components: estimates of the amount of water people ingest in a given time period (ingestion rate) and estimates of how long people spend in the recreational activity (duration). There are no New Zealand data on water ingestion rates or duration of activities at freshwater sites. As such, data from overseas studies of countries with similar bathing season climates to those observed in New Zealand were used as a surrogate for New Zealand data.

11.7.1 Ingestion rate

The rate water is ingested during freshwater swimming activities is highly variable between individuals and repeated activities by individuals. Hence, the ingestion rate is represented in the model as a distribution.

Dufour et al (2017) conducted a study in chlorinated open-air swimming pools. Participants were asked to swim for approximately one hour and then urine was collected for a period of

[6] https://sportnz.org.nz/resources/insights-tool/ (Accessed 8 November 2023)

24 hours following the swim. The quantity of a chemical marker (cyanuric acid[7]) in the urine and the pool water allowed an estimate of water ingestion to be made.

The raw data from the Dufour study was used to generate a distribution of water consumption rates (ml/h) for the 6 – 12 year age group. The data were fitted to a Weibull distribution with parameters: shape = 1.15, scale = 57, and maximum value of 245 (Appendix B.5.1).

11.7.2 Duration of swimming activity

The duration of swimming activity in freshwater is variable between people and for the same person on different visits to a site. The duration is dependent on a number of environmental factors (air and water temperature, river flow and depth) as well as the personal preference of the swimmer. Swim duration is likely to differ with geographical location in New Zealand.

A pooled analysis of four summer epidemiological studies at freshwater temperate beaches in the USA found children in the 4 – 12 year age group spent a median of one hour and maximum of six hours in the water during a beach day (DeFlorio-Barker et al, 2018). A Dutch study of freshwater swimming activities, suggested a median duration for those aged under 15 of one hour, and a 95th percentile estimate of five hours (Schets et al, 2011).

The QMRA uses a distribution fitted to the summary statistics for 4 to 12-year-olds in the study by (DeFlorio-Barker et al, 2018). Hours of swimming activity were represented by a truncated gamma distribution (shape = 1.8, rate = 0.02, min = 2, max = 360). This distribution allows for inclusion of occasional longer duration events in the model.

11.8 SIMULATED VOLUME OF WATER INGESTED

The simulation model includes the variability in water ingestion rate and swim duration to give a resulting distribution of total water ingested during a swimming event. The distributions and summary statistics for the open-air swimming ingestion rate, the duration of swimming and the simulated volume of water consumed are provided in Figure 30. Half of swimmers ingest volumes of water < 45 ml and 95% of swimmers ingest volumes < 275 ml. The higher volumes of ingested water theoretically possible in the model are there to capture rare events of people ingesting higher volumes.

International recreational water QMRAs have used reference values for the volume of water unintentionally ingested while swimming in the range 20 – 100 ml (World Health Organization, 2016). Other QMRA have used distributions based on the results from the Dufour study combined with fixed durations in the range of 30 – 60 minutes (Skiendzielewski et al, 2024; United States Environmental Protection Agency, 2024) or combined with durations observed by DeFlorio-Barker et al (2018) and Boehm and Soller (2020). The sensitivity of the simulation model outputs to the choice of representation of the volume of water ingested is discussed in Appendix G.

[7] Cyanuric is added to swimming pool water to stabilise chlorine. It is not metabolised by humans and is excreted quantitatively in urine.

11.8.1 Assumptions

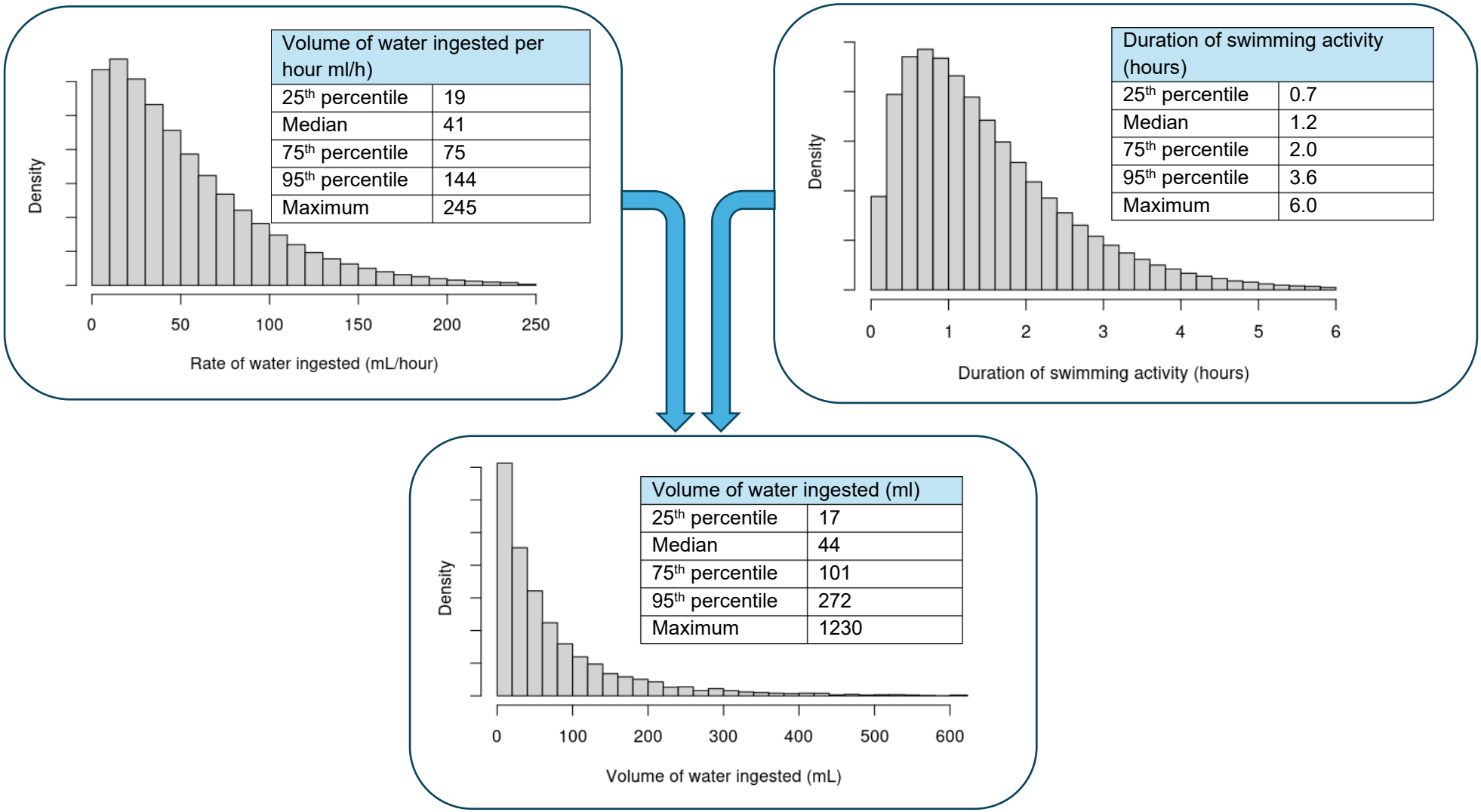
1. *The distributions for water ingestion and duration of swimming are appropriate for New Zealand children.*

   There is no New Zealand data to compare against or to inform on the possible uncertainty in the distribution parameters.

2. *The water ingestion rate is independent of the swimming duration*.

   No data was found to inform this assumption. A sensitivity analysis of the probability of infection if increased water ingestion rates were associated with shorter durations of swimming activity has been performed (Appendix D).

Figure 30: Model distributions for the volume of water ingested by 6 - 12-year-olds during open-air swimming

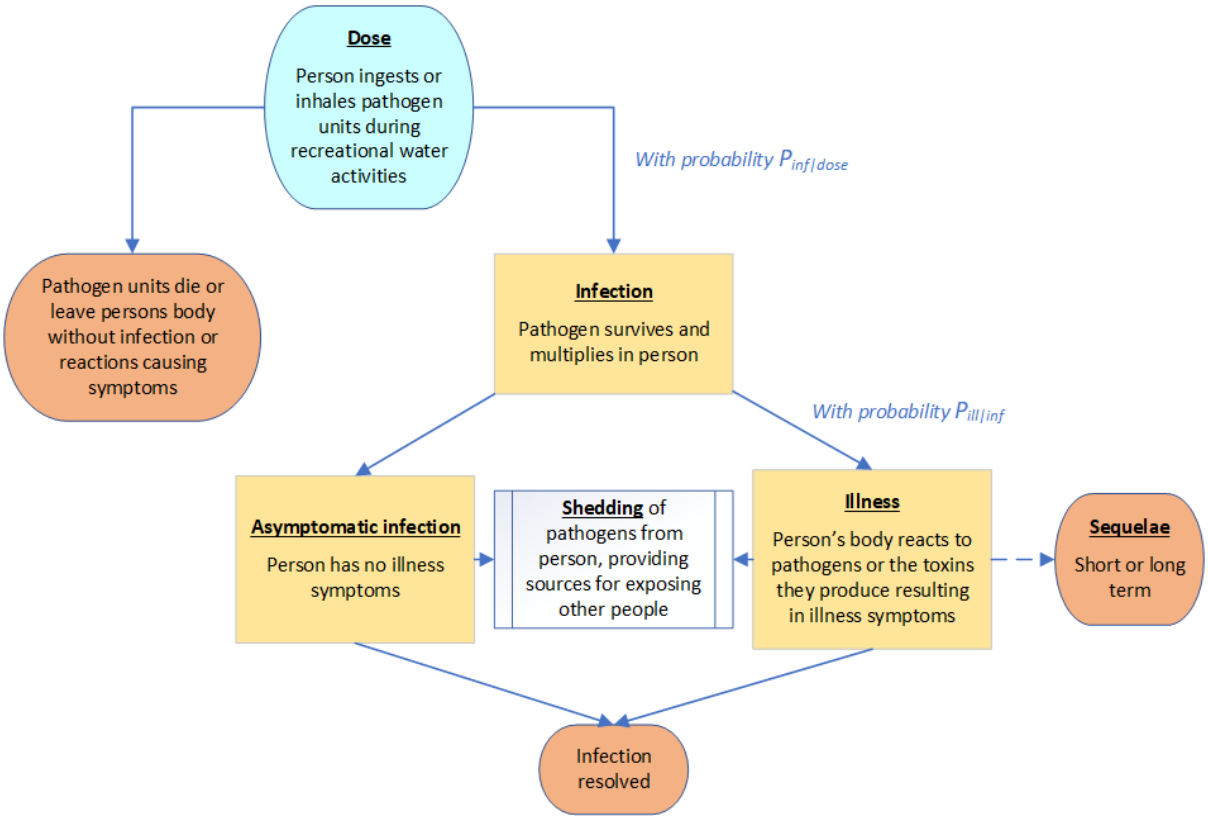


12 HEALTH EFFECTS ASSESSMENT

12.1 INTRODUCTION

If a person ingests or inhales pathogenic micro-organisms from exposures during recreational water activities, there are a number of possible outcomes that can occur as shown in Figure 31. The person may become infected, which occurs when pathogen units multiply in body tissues or organs. An infected person may go on to exhibit symptoms (becomes ill) or be asymptomatic (have no symptoms). Dose-response models are used to define the probability of a person becoming infected given they have been exposed to a given dose of pathogen units ($P_{inf \mid dose}$). Models are also used to describe the probability of a person having illness symptoms given they are infected ($P_{ill \mid inf}$).

Figure 31: Health effects flow diagram



Infected people may shed pathogens via faecal matter, vomit, urine, oral or nasal fluids, irrespective of whether they have illness symptoms. For the pathogens of interest to this project, the majority of people who become ill will recover, often without treatment (self-limiting). However, in a small proportion of cases, long term illness (sequelae) or even death may result from the initial illness.

12.2 HEALTH CONSEQUENCES

Table 45 provides a summary of the possible health consequences of infection with the pathogens considered in this simulation model as well as for norovirus, a pathogen commonly considered in recreational water QMRAs and considered further in section 14. More detailed information about the health effects of these pathogens with relevant reference links are provided in Appendix A.

12.3 DOSE-RESPONSE

This QMRA focuses on exposure to selected pathogens during swimming at freshwater sites. During freshwater swimming, ingestion of pathogens via swallowing water is likely to result in greater exposure than cross-contamination via the hands to the mouth. Dose-response relationships from the literature relating to oral exposure were reviewed for use in this QMRA.

Dose-response relationships includes elements of both variability and uncertainty. There is variability in:

- the infectivity of different strains of a pathogen
- the immune response people have to a pathogen dose at a given point in time
- the dose ingested, given irregularly dispersed pathogens in the water

Dose-response models have been developed from data from trials or outbreaks with limited ranges of pathogen strains and often involve people from specific demographic groups. Therefore, there is uncertainty in the applicability of dose-response models to other population groups or for other strains of the pathogen. For outbreak data there is uncertainty in the dose of pathogen ingested.

With ethical barriers to the collection of new trial data to inform dose-response relationships, the current models provide estimates of the dose-response relationships based on data available at this time. The dose-response relationships do not allow for recognition of dose-response differences for more susceptible populations, such as immune compromised people.

Table 46 summarises the dose-response models for infection and illness used in this study. Further details of the dose-response relationships can be found in Appendix F. Appendix F also compares the dose response simulation results with the results of two-dimensional dose response relationships which also incorporate uncertainty in the dose-response parameters (T. Burch, 2020; Teunis et al, 2018).

Table 45: Health effects of infection with pathogens considered in this study and norovirus

Microbe type	Pathogen	Mode of exposure	Common symptoms	Usual duration of symptoms	Recognised sequelae	Notified cases hospitalised / Total notified cases (%) in 2023	DALY per case
Bacteria	*Campylobacter* spp.	Ingestion	Muscle pain, headache, fever, diarrhoea, abdominal pain and nausea	3 – 7 days	GBS, ReA, IBS	989/6089 (16%)	0.015
	Salmonella spp.	Ingestion	Diarrhoea, abdominal cramps, nausea, vomiting	1 – 7 days	ReA, IBS	218/827 (26%)	0.018
	STEC	Ingestion	Diarrhoea, abdominal cramps, fever, vomiting, thrombocytopaenia[a], uraemia[b]	5 – 7 days	HUS with possible progression to ESRD	235/1006 (23.4%)	0.13
Protozoa	*Cryptosporidium* spp.	Ingestion	Smelly watery diarrhoea, abdominal cramps, nausea, vomiting, fever	2 weeks to a month	Persistent diarrhoea, abdominal pain, nausea, fatigue or headaches	67/831 (8.1%)	0.012
	Giardia spp.	Ingestion	Abdominal cramps, nausea, acute or chronic foul-smelling diarrhoea, low-grade fever, anorexia and bloating	7 – 10 days, can last months if untreated	IBS (possibly making existing mild IBS worse) Chronic fatigue. [Sequelae not included in DALY calculations]	43/897 (5%)	0.011
Virus	Norovirus	Inhalation or ingestion	Nausea, vomiting, abdominal cramps, diarrhoea,	0.5 – 3 days	None	Data unavailable – only subset of individual cases notified	0.01

[a] Low level of platelets in the blood;[b] Build-up of waste products in the blood due to impaired kidney function. DALY, disability-adjusted life year; GBS, Guillain-Barré syndrome; ReA, reactive arthritis; IBS, irritable bowel syndrome; HUS, haemolytic uraemic syndrome; ESRD. End-stage renal disease

Table 46: Dose response relationships used in the Monte Carlo simulation model considering variability in response to ingestion of the pathogen

Pathogen	Probability of infection given dose (*D*) ingested		Probability of illness given infection has occurred	
	Model	**Reference**	**Model**	**Reference**
Campylobacter jejuni	$P_{\inf \mid D} = 1 - \left(1 + \frac{D}{\beta}\right)^{-\alpha}$ $\alpha = 0.145, \quad \beta = 7.59$	Medema et al (1996)	$Uniform(\min = 0.2, \max = 0.6)$	Based on doses in the order of 10^2 to 10^4 from Black et al (1988) Unpublished dataset from (Teunis et al, 2018)
***Cryptosporidium* spp.**	$P_{\inf \mid D} = 1 - e^{-rD}$ $r = 0.09$	United States Environmental Protection Agency (2024)	$Uniform(\min = 0.3, \max = 0.7)$	Jahne et al (2024) United States Environmental Protection Agency (2005)
Giardia lamblia	$P_{\inf \mid D} = 1 - e^{-rD}$ $r = 0.0199$	Rose et al (1991)	$Uniform(\min = 0.2, \max = 0.7)$	Eisenberg et al (1996) Jahne et al (2024)

Refer to Appendix E for further information on these dose response models.

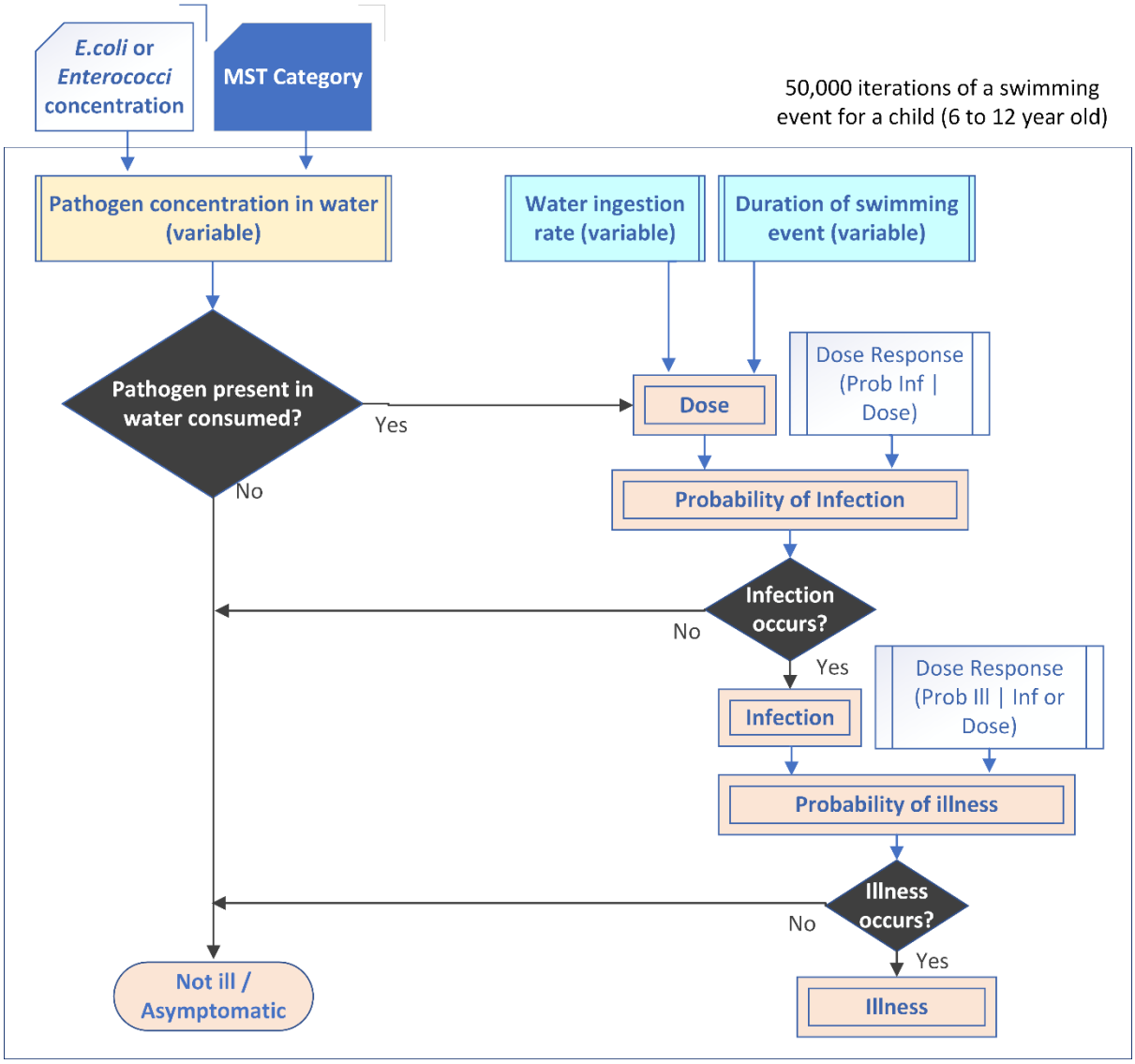
13 RISK CHARACTERISATION

13.1 SIMULATION APPROACH

Monte Carlo simulations were conducted to estimate pathogen exposure, combine the exposure estimate with a dose-response relationship to give a probability of infection or illness, and to characterise risks in terms of an estimate of the percentage of people becoming infected or becoming ill from exposure to *Campylobacter*, *Giardia* or *Cryptosporidium* during swimming at freshwater recreational sites (Figure 32).

The simulation model is designed as a high-level screening model for assessing risk to inform national guidelines. The model inputs are *E. coli* or enterococci concentrations and MST sources for a water sample. Site or time specific information such as rainfall are not included in the model inputs. The survey data used in the model covers a range of sites and environmental conditions.

Figure 32: Monte Carlo simulation model



Variable components of the model include:

- the volume of water ingested by swimmers (6-12 year old age group), due to variability in the amount of water ingested during, and the duration of a swimming event at a freshwater site
- the distribution of pathogen concentrations for given levels of indicators in the water samples collected by the study
- people's infection and illness response to a dose of pathogens.

For each scenario, the simulation model was run for 50,000 iterations to ensure convergence of the risk metric (illness occurrence percentage estimates to within 0.2%). Each iteration represents a person taking part in swimming at a freshwater site. The dose, probability of infection, if infection occurred, the probability of illness and if illness occurred was recorded for each iteration.

13.2 CAMPYLOBACTERIOSIS RISK

This section provides the results from the simulation model either using the *Campylobacter* concentration distribution (section 11.3) based on the data collected all year round, or data that was collected during the bathing season. For the purpose of reporting at a national level the bathing season is defined as 31 October to 31 March.

13.2.1 Campylobacteriosis by *E. coli* concentration

The estimated number of people becoming infected or ill with *Campylobacter* increases monotonically for swimming in water with increasing *E. coli* concentrations (Table 47 and Figure 33). Risks, expressed as the probability of an individual becoming infected or the proportion of people becoming infected, increase from 2% at a concentration of 50 *E. coli*/100 ml to 10% at concentrations greater than 2000 *E. coli*/100 ml. Restricting the model data to that collected during the bathing season, decreases the risk estimates to 1% at 50 *E. coli*/100 ml and 8% at concentrations greater than 2000 *E. coli*/100 ml.

The model outputs (Table 47) and the prediction of non-zero *Campylobacter* concentrations at low *E. coli* concentrations (Figure 17, section 9.1) suggest the concept of a zero risk of campylobacteriosis while conducting swimming type activities at freshwater sites may not be realistic. At an *E. coli* concentration of 50 *E. coli*/100 ml it is estimated some but less than 0.5% (or less than five in 1000) of summer swimmers would become ill from *Campylobacter*.

The last three rows of Table 47, provide the model outputs for the *E. coli* concentrations (130, 260, 540 *E. coli*/100 ml) used in the National Policy Statement for Freshwater Management 2020, Table 22: *E. coli* for primary contact sites during the bathing season (New Zealand Government, 2020). The infection occurrence estimates from the 2025 model are higher than generated from the 2002 model both using all data and using bathing season data (McBride et al, 2002). The differences in the risk estimates and models are discussed in section 15.

The estimated DALY rates per 1000 swims are also provided in Table 47, based on previous estimates of DALYs per case. These provide a measure of the severity of disease as well as frequency of occurrence, which together provide a measure of the burden of disease (WHO, 2016). One DALY represents the loss of the equivalent of one year of full health. For the bathing season model the DALY estimates range from 0.1 to 0.5 DALY per 1000 swim events.

Table 47: Estimated percentage of swims in water with different *E. coli* concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. Based on data collected all year or during the bathing season.

E. coli / 100 ml	Log_{10} *E. coli* /100 ml	Percentage of swims resulting in infection		Percentage of swims resulting in illness		DALY per 1000 swims	
		All year data	Bathing season data	All year data	Bathing season data	All year data	Bathing season data
50	1.7	2.0	1.0	1.0	< 0.5	0.1	< 0.1
100	2.0	3.0	1.5	1.0	0.5	0.2	0.1
200	2.3	4.5	3.0	2.0	1.0	0.3	0.2
300	2.5	6.0	4.5	2.5	2.0	0.4	0.3
400	2.6	7.0	6.0	3.0	2.5	0.4	0.4
500	2.7	8.0	7.0	3.0	2.5	0.5	0.4
600	2.8	9.0	7.5	3.5	3.0	0.5	0.5
800	2.9	9.5	8.5	4.0	3.5	0.6	0.5
1000	3.0	9.5	9.0	4.0	3.5	0.6	0.5
> 2000	> 3.3	10.5	8.0	4.5	3.5	0.6	0.5
130	2.1	[0.0][a] 3.5	[0.0] 2.0	1.5	0.5	0.2	0.1
260	2.4	[0.9] 5.5	[0.4] 4.0	2.0	1.5	0.3	0.2
540	2.7	[4.9] 8.5	[2.5] 7.0	3.5	3.0	0.5	0.5

Note: a: [ ] Percentage of swims resulting in *Campylobacter* infection as estimated by the 2002 model (McBride et al., 2002). In consideration of the modelling uncertainties, the percentage of swims estimated to result in infection are rounded to nearest multiple of 0.5%, and DALY estimates are rounded to one decimal place.

Each iteration of the simulation model represents a single scenario of a person undertaking a swimming activity. The percentage of swimmers becoming ill or infected over the 50,000 model iterations provides the occurrence percentages in Table 47. The percentage of people becoming ill after swimming in water with a given *E. coli* concentration is also plotted in Figure 33(a).

Across the scenarios represented by the model iterations, some swimmers will have a higher probability of infection and subsequent illness than others, due to the variability in the

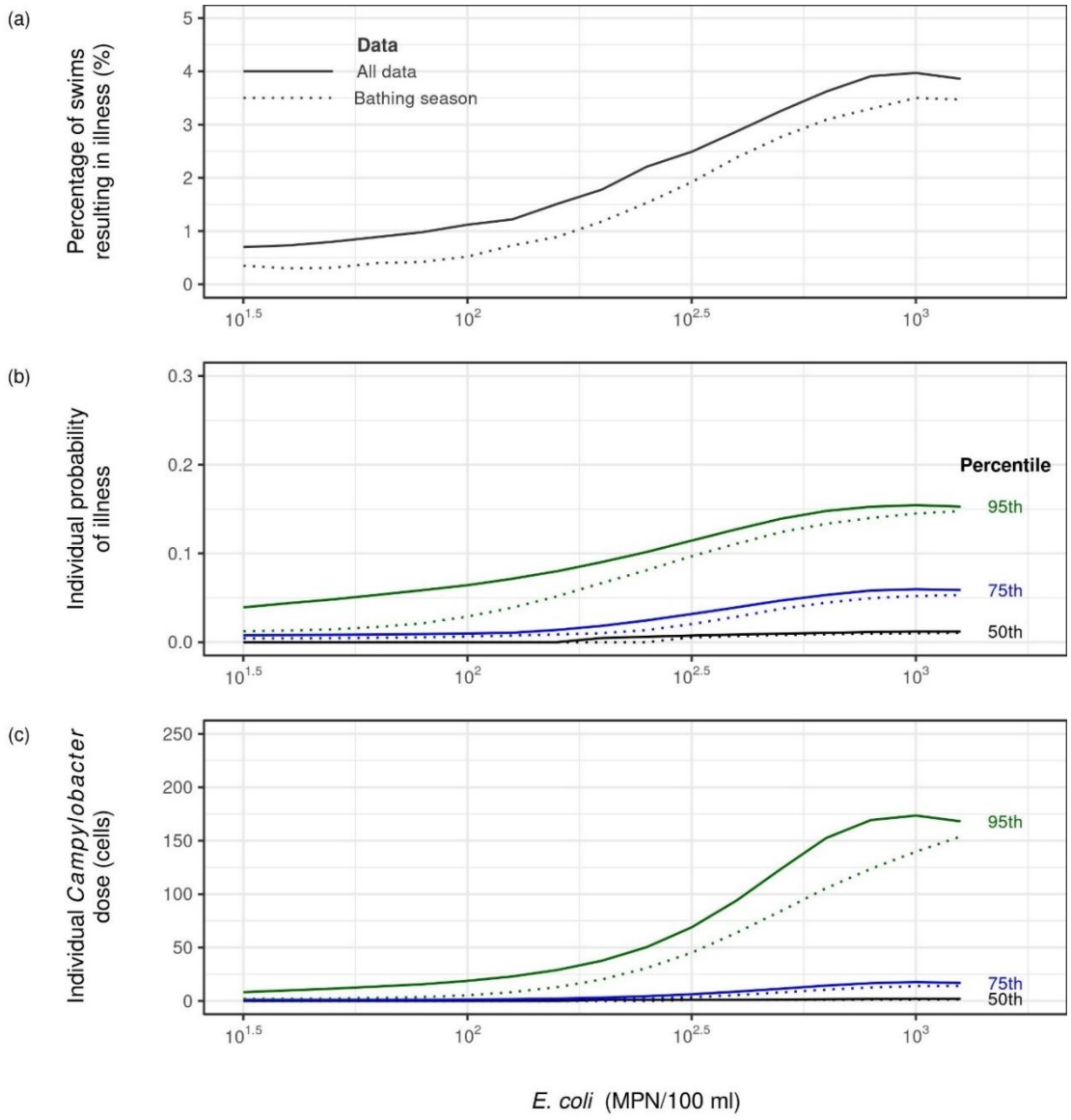
amount of water consumed, the concentration of *Campylobacter* concentration in the water ingested and the response of the individual to the dose of cells consumed. The 50th, 75th and 95th percentiles of the distribution of individual probability of illness and individual doses are provided in Figure 33 (b and c).

For example, considering the dose and probability of illness estimates for scenarios using survey data from the bathing season only, at a concentration of 500 *E. coli* /100 ml ($10^{2.7}$), half of the swimmers ingested no more than a single *Campylobacter* cell, and 5% ingested more than 83 cells (Figure 33c). At the same *E. coli* concentration, the individual probability[8] of campylobacteriosis was below 0.008 (equivalent to 0.8%) for half of the swimmers and was above 0.12 (equivalent to 12%) for 5% of swimmers (Figure 33b).

The illness occurrence percentage and individual probabilities of illness both increase with increasing *E. coli* concentration in the swimming water, until *E. coli* concentrations reach about 800 MPN/100 ml, after which there is minimal increase in risk.

[8] A probability is the chance that an event will occur and ranges in values between zero for no possibility of it happening and a probability of one when it will definitely happen. A probability of 0.5 means it is equally likely to happen as not.

Figure 33: Estimated *Campylobacter* illness occurrence and 50th, 75th and 95th percentiles of individual probability of illness and *Campylobacter* dose in water with different *E. coli* concentrations. Data collected all year round or collected during the bathing season.



13.2.2 Campylobacteriosis by enterococci concentration

The simulation model was run for enterococci concentrations in the range of 40 to 1000 MPN/100 ml. There were insufficient water samples with enterococci concentrations above 1,000 MPN/100 ml to support risk assessment.

The estimated number of people becoming infected or ill with *Campylobacter* increased monotonically for swimming in water with increasing enterococci concentrations (Table 48 and Figure 34), from 4.5% infection rate at 40 enterococci/100 ml to 12.5% infection rate at 1000 enterococci/100 ml. Restricting the model data to that collected during the bathing season, the infection rate was lower at 2.5% infection rate at 40 enterococci/100 ml to 12% infection rate at 1000 enterococci/100 ml. As with risk estimates based on *E. coli* concentrations, enterococci concentrations result in non-zero estimates of the risk of campylobacteriosis at all non-zero enterococci concentrations.

The estimated DALYs per 1000 swims are also provided in Table 48. For the bathing season model the DALYs range from 0.2 to 0.7 DALYs per 1000 swim events for enterococci concentrations in the range from 40 to 1000 MPN/100 ml.

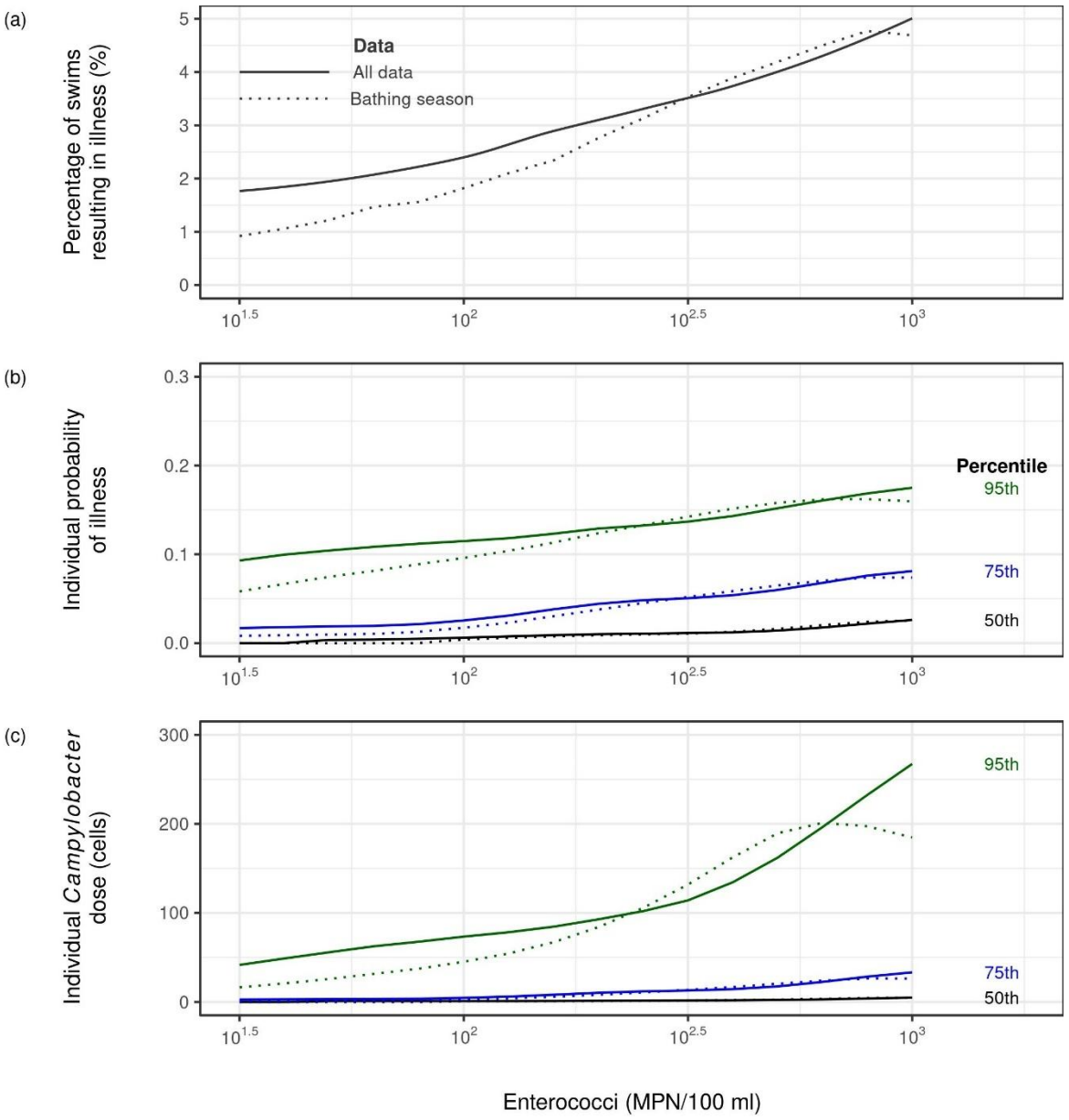
Table 48: Estimated percentage of swims in water with different enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1000 swims. Based on data collected all year or during the bathing season.

Enterococci / 100 ml	Log_{10} enterococci /100 ml	Percentage of swims resulting in infection		Percentage of swims resulting in illness		DALY per 1000 swims	
		All year data	Bathing season data	All year data	Bathing season data	All year data	Bathing season data
40	1.6	4.5	2.5	2.0	1.0	0.3	0.2
100	2.0	6.0	4.5	2.5	2.0	0.4	0.3
200	2.3	7.5	7.0	3.0	3.0	0.5	0.5
300	2.5	8.5	8.0	3.5	3.5	0.5	0.5
400	2.6	9.0	9.5	3.5	4.0	0.5	0.6
500	2.7	9.5	10.5	4.0	4.5	0.6	0.6
600	2.8	10.5	11.0	4.0	4.5	0.6	0.7
800	2.9	12.0	12.0	5.0	5.0	0.7	0.7
1000	3.0	12.5	12.0	5.0	5.0	0.8	0.7

Note: In consideration of the modelling uncertainties, the percentage of swims estimated to result in infection are rounded to nearest multiple of 0.5% and DALY estimates are rounded to one decimal place.

Comparison of the model outputs from simulations using data collected all year round with outputs based on data from the bathing season only (Table 48 and Figure 34) shows there is not a consistent difference in risk estimates. For lower enterococci concentrations the percentage of swimming events resulting in illness is less for simulations based on the bathing season than for simulations using data from the whole year. However, for higher enterococci concentrations the results are similar for the two data sets.

Figure 34: Estimated *Campylobacter* illness occurrence and 50th, 75th and 95th percentiles of individual probability of illness and *Campylobacter* dose in water with different enterococci concentrations. Data collected all year round or collected during the bathing season.



13.2.3 Sensitivity and uncertainty analyses

Water ingestion

The sensitivity analysis of the simulation model outputs to the choice of representation of the volume of water ingested is provided in Appendix F.3. The model is sensitive to the choice of volume of water ingested, which would be expected given the dose is directly related to the volume of water ingested. A fixed volume of ingested water of 50 ml results in similar estimates of swimmer campylobacteriosis infection rate as produced by the model with the 6-12 year old variable ingestion volume.

To determine if infection occurence changes if increased water ingestion rates were associated with shorter durations of swimming activity, a sensitivity anlysis was performed (appendix F.3.2). The analysis estimates a decrease of between 0.4 and 0.9 in the percentage of swimmers being infected if swimmers who swallow more than 50 mL/h, have shorter swim durations as the rate of water ingestion increases.

Dose-Response

The 6-12 year old age group was chosen as a conservative approach to risk assessment, as they are likely to ingest more water during a visit to a freshwater site than other age groups. However, the available dose-response relationships are not age specific, and are mainly based on adult feeding or vaccine trials.

The feeding trials have limited low dose data, and no doses below 100 cells. There are no data to directly inform dose-response relationships for the low doses seen in the majority of the simulation iterations (Figure 33). Another dose-response relationship that utilises data from a wider range of sources than the data used in the base model, but still no known doses below 100 cells, has been proposed by Teunis et al. (2018). This model also incorporates uncertainty in the dose-response parameters (appendix E.1).

Using the Teunis et al. (2018) dose-response approach in the simulation model estimates illness rates that are double those generated by the Medema et al. (1996) dose-response relationship. The estimated uncertainty interval for the percentage of swimming events resulting in campylobacteriosis is < 0.5% to 13.5% for swimming in water with 100 *E. coli*/100 ml, and 0.5% to 35% for swimming in water with 1,000 *E. coli*/100 ml. The wide uncertainty intervals reflect the uncertainty in the model parameters at low doses.

While the dose-response relationship represents the largest source of uncertainty in the model outputs, using a consistent dose-response model allows the effects of other variables to be identified. For this reason, the campylobacteriosis dose-response used in the 2002 QMRA model (Medema et al, 1996) has been used for the 2024 simulation model. If further dose-response or epidemiology data becomes available in the future, the dose-response aspect of the simulation model should be re-evaluated.

13.3 CRYPTOSPORIDIUM AND GIARDIA RISK

13.3.1 Simulation results

The estimated percentage of swimming events resulting in illness and/or infection from *Cryptosporidium* or *Giardia* are all below one percent in the range of enterococci concentrations found in the freshwater survey (Table 49). For the water samples collected in this study, the rate of infection and subsequent illness for giardiasis and cryptosporidiosis is less than estimated for campylobacteriosis.

Enterococci and *E. coli* are less useful indicators for *Cryptosporidium* or *Giardia*. Enterococci and *E. coli* (similar results not presented) only show a minor increasing trend of infection and illness rates with increasing concentrations (Table 49).

The slightly higher infection percentage for *Cryptosporidium* than *Giardia* relates to the rare but higher concentrations in the water, which are slightly more frequent for *Cryptosporidium* than *Giardia*. Any future modelling of protozoa in rivers may be better to focus on the rare events rather than treating the concentrations as a continuous potential occurrence.

Table 49: Estimated percentage of swimming in water with different enterococci concentrations resulting in *Cryptosporidium* or *Giardia* infection, illness and DALYs per 1000 swims.

Enterococci / 100 ml	*Cryptosporidium*			*Giardia*		
	Infection (percentage of swims)	Illness (percentage of swims)	DALY per 1000 swims	Infection (percentage of swims)	Illness (percentage of swims)	DALY per 1000 swims
100	0.5	0.5	< 0.1	< 0.5	< 0.5	< 0.1
1000	1.0	0.5	0.1	0.5	< 0.5	< 0.1

Note: In consideration of the modelling uncertainties, the percentage of swims estimated to result in infection are rounded to nearest multiple of 0.5%, and DALY estimates are rounded to one decimal place.

13.3.2 Uncertainty in simulation estimates

Analysis bias

The water samples collected in this study, were from undisturbed water 200 mm below the surface. Some samples in turbid water, were not analysed due to the filters being unable to filter sufficient water for analysis. As protozoa can be associated with particles and sediments that may be present in streams (Atherholt et al, 1998), it is possible higher protozoa concentrations would have been identified, if these samples had been analysed.

To consider the potential bias in the results from excluding some samples with low volumes of water filtered, the *Cryptosporidium* model was run with imputed values for the missing *Cryptosporidium* concentration values. The imputed values were in the range of 2 to 140 oocysts/10 L, which is the range of values seen in low water volume samples where analyses were completed (appendix F.4). Imputing for the missing values increased the percentage of swims resulting in infection by 0.1 to 0.4% and percentage of swims resulting in illness by 0 to 0.2%.

The exclusion from analysis of some low water volume samples, may have led to a small underestimate in the percentage of swimmers becoming ill and/or infected. However, the estimates for infection and illness rates remain well below those for *Campylobacter*.

Dose response

Appendices E.2 and E.3 compare the simulation outputs for the base model with alternative published dose-response models for cryptosporidiosis and giardiasis.

The fractional-Poisson model (M. J. Messner and P. Berger, 2016) for cryptosporidiosis assumes that 74% of people will be highly susceptible to infection, and 26% perfectly immune. For susceptible people this means doses of 5 oocysts or more will result in infection, and even 1 oocyst would result in infection 63% of the time. This rate of infection for low doses produced by the simulation model is greater than for the exponential model used in the base simulations, resulting in a higher estimate of infection rates by a factor of about 5.

Appendix E.3 provides the *Giardia* model results using a dose-response approach proposed by Burch (T. R. Burch, 2020) which incorporated uncertainty in the dose-response parameters. The resulting uncertainty intervals were < 0.1 to 0.9% of swimming events resulting in illness for the water with 100 enterococci /100 ml, and < 0.1 to 2.4% swims resulting in illness for water with 1,000 enterococci /100 ml.

The illness rates estimated by the alternative protozoa dose-response approaches are below the rates estimated by the base simulation model for campylobacteriosis.

13.4 RISK CHARACTERISATION CONCLUSIONS

The simulation model estimates the illness and/or infection risk for swimming at freshwater sites, accounting for the variability of swimmer behaviour and concentrations of indicators and pathogens in the water.

The model uses the relationships between faecal indicator bacteria and pathogens based on the data collected during the 2020 to 2024 freshwater survey. The simulation model outputs provide the following information:

- *E. coli* or enterococci can be used to define changes in the percentage of swimming events resulting in campylobacteriosis, but are less suitable for use for risks of cryptosporidiosis or giardiasis.
- Estimated *Campylobacter* concentrations result in a higher estimated percentage of swimmers becoming ill from swimming in water with given *E. coli* or enterococci concentrations, than for *Cryptosporidium* or *Giardia* concentrations predicted from the same *E. coli* or enterococci concentrations. This holds true even after consideration of uncertainty in the dose-response relationships and analysis bias for the protozoa.
- Even at low *E. coli* (50 MPN/100 ml) or enterococci (40 MPN/100ml) concentrations, there can be a non-zero risk of campylobacteriosis (1-2% of swims resulting in illness).
- The relationship between the concentrations of *E. coli* and *Campylobacter* is slightly different when considering data collected during 31st October to 31st May (bathing season as defined for study), compared to data collected all year round. The difference is sufficient to change the percentage of swimming events resulting in infection to 0.5 to 1.5% lower in the bathing season than using data collected all through the year.
- The percentage of swimming events resulting in infection for different *E. coli* concentrations is predicted to be slightly higher than predicted by the 2002 QMRA model. The differences in the two modelling approaches will be discussed in section 15.2.

14 ESTIMATING RISK OF NOROVIRUS INFECTION ON THE BASIS OF CRASSPHAGE LEVELS

14.1 INTRODUCTION

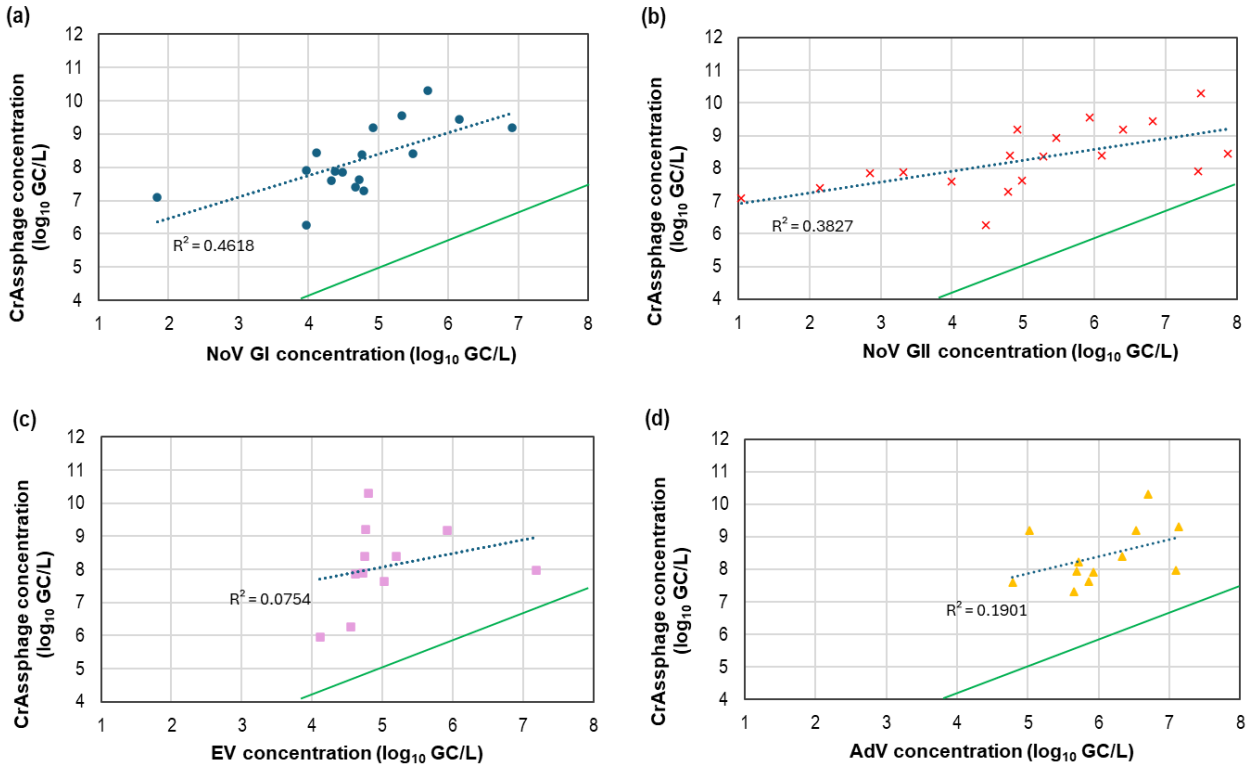
The previous analysis is based directly on the detection and measurement of *Campylobacter*, *Cryptosporidium* and *Giardia* in river water samples. However human viruses are also a known risk for freshwater impacted by human faecal matter. In this section we describe an approach for estimating of the risk from norovirus based on estimates of the relative levels of CrAssphage to norovirus as measured in human wastewater.

QMRA analyses have been undertaken, based on the levels of MST markers in sewage relative to human norovirus (Ahmed, Korajkic, et al, 2024; Boehm et al, 2015; Crank et al, 2019; Schoen et al, 2020). These studies established MST marker concentrations that correspond to specified gastrointestinal illness rates termed a risk-based threshold (RBT). Initial work focussed on use of HF183 and other bacteria markers (Boehm et al, 2015). While useful, as highlighted in this study, HF183 has sources that are not sewage associated in New Zealand (Devane et al, 2013), and as a bacterial marker are less applicable to viruses

Other research has made use of viral indicators, including CrAssphage (Ahmed, Korajkic, et al, 2024; Crank et al, 2019; Schoen et al, 2020), to derive RBT for human viruses. CrAssphage is a viral marker highly specific for human faecal contamination and has been widely used in studies of microbial abundance in water (Mafumo et al, 2023). We reviewed 33 published papers that assessed the concentration of CrAssphage, norovirus, and other pathogenic viruses in wastewater collected across the world.

Regardless of the laboratory techniques used, CrAssphage was always present in higher abundance than pathogenic viruses. The concentration of CrAssphage was also indicative of the level of norovirus pathogen contamination. In wastewater influent, crAssphage and norovirus concentrations were correlated (R^2 0.38-0.46), but CrAssphage was present at 2.5 – 5.5 $\log_{10}$ copies/L per sample higher than norovirus genogroup GI (Figure 35 a), and 0.5 – 6 log10 copies/L per sample higher than norovirus genogroup GII (Figure 35 b). CrAssphage concentration was less correlated with enterovirus (R^2 0.08) or human adenovirus (R^2 0.19) in wastewater influent but was still always present at higher concentrations (Figure 35, c and d).

Figure 35: Linear correlation (dashed blue line) between CrAssphage concentration in wastewater influent and: (a) norovirus genogroup GI (NoV GI); (b) norovirus genogroup GII (NoV GII); (c) enterovirus (EV); and (d) human adenovirus (AdV) concentration. The diagonal green line is the equivalence line upon which samples would fall if the same concentration of both viruses was found in the sample.



14.2 USE OF CRASSPHAGE LEVELS TO ESTIMATE THE RELATIVE LEVEL OF NOROVIRUS

This section provides an example of how a QMRA simulation model could be used to estimate the risk of norovirus infection and illness given knowledge of crAssphage concentration in the water. The simulation model follows these steps for each iteration representing a swim in water with a given CrAssphage concentration:

1. Estimate the norovirus concentration for the target CrAssphage concentration.
2. Calculate the dose of norovirus by the swimmer given sampling from the same water ingestion distribution as the *Campylobacter* simulation model (section 11.8).
3. Apply a norovirus dose response, to predict if the swimmer will become ill or infected. The dose response applied to this model was a model for the probability of infection given dose, and a fixed 0.6 probability of illness given infection. The dose response also assumed 74% of the population is susceptible to norovirus infection and applied a dose harmonisation factor (as implemented in (Cressey, 2021)).

The results from 50,000 simulation iterations were combined to estimate the percentage of people swimming in water with a given crAssphage concentration becoming infected or ill from norovirus.

14.2.1 Estimating the norovirus concentration in water at swimming site

The estimate of norovirus concentration at the swimming site is based on the following assumptions:

- There is a consistent relationship between the concentrations of norovirus and CrAssphage in bulk human untreated sewage (Ahmed, Korajkic, et al, 2024; Crank et al, 2019; Schoen et al, 2020).
- The addition of faecal matter to the river is recent enough to the swimming event that no die off occurs.
- The norovirus and CrAssphage will be diluted by the same factor, given they come from the same faecal source.

The concentration of norovirus at a recreational water site can then by estimated by the paired equations:

$$C_{noro\_rec} = C_{noro\_sewage} \times \text{Dilution factor} \times \text{norovirus die off factor} \quad [14.1]$$

$$C_{crA\_rec} = C_{crA\_sewage} \times \text{Dilution factor} \times \text{crAssphage die off factor}$$

C_{noro_rec} is the estimated concentration of norovirus in the recreational water, C_{crA_rec} is the measured concentration of CrAssphage in the recreational water, C_{noro_sewage} and C_{crA_sewage} are the concentrations of norovirus and CrAssphage in sewage. The die off factor is set to one representing fresh faecal matter, where no die off has occurred.

Combining these two equations and applying the above assumptions, the concentration of norovirus in the water at the swim site can be estimated by:

$$C_{noro\_rec} = C_{noro\_sewage} \times \frac{C_{crA\_rec}}{C_{crA\_sewage}} \quad [14.2]$$

To estimate the concentration of norovirus in water at the swimming site for each iteration in the simulation, the $\log_{10}$ distributions for norovirus and CrAssphage in raw sewage were modelled by a log normal distribution (mean=5.0, sd=0.9) and Weibull distribution (shape=31.6 and scale=9.4) respectively (Ahmed et al., 2024). The sampling from these concentrations reflected a ranked correlation between norovirus and CrAssphage of 0.4.

14.2.2 Simulation model results

The estimated percentage of people becoming infected with norovirus after swimming in water of different CrAssphage concentrations in shown in Table 50, along with the number of water samples from the survey within different ranges of CrAssphage concentrations. Water samples associated with aged human faecal matter are excluded in the counts, as the model is based on fresh faecal contamination.

CrAssphage concentrations of 500, 5000, 30,000 and 90,000 GC/100 ml corresponded to an estimated 0.1%, 1%, 5% and 10% of swimmers becoming infected from norovirus. Illness risk corresponds to <0.1%, 0.6%, 3.0%, and 6.0%.

Table 50: Risk of infection calculated from copies of CrAssphage, and number of samples in each category (excluding samples identified by MST as including aged human faecal contamination).

CrAssphage GC/100 ml	Percentage of swimmers becoming infected	Percentage of swimmers becoming ill
500	0.1 %	< 0.1 %
2,500	0.5 %	0.3 %
4,500	1.0 %	0.6 %
10,000	2.0 %	1.2 %
20,000	3.7 %	2.2 %
30,000	5.1 %	3.0 %
60,000	8.0 %	4.8 %
90,000	10.0 %	6.0 %
> 90,000	>10.0 %	>6.0 %

CrAssphage interval GC/100 ml	Samples in survey (N=916)
≤ 500	652 (71.2 %)
501 to 2,500	61 (6.7 %)
2,501 to 4,500	27 (2.9 %)
4,501 to 10,000	35 (3.8 %)
10,001 to 20,000	29 (3.2 %)
20,001 to 30,000	15 (1.6 %)
30,001 to 60,000	24 (2.6 %)
60,001 to 90,000	24 (2.6 %)
> 90,000	49 (5.3 %)

Applying this model to samples in this study, would indicate that approximately 8% of the survey water samples correspond to exceeding five percent of swims resulting in norovirus illness (Table 50). These samples were from eight different rivers. All eight of the rivers had high levels of *E. coli* (95th percentile > 6,000).

Ahmed et al (2024) found that CrAssphage had a slightly higher die off rate than norovirus in water in estuarine waters, both in daylight and the dark. After one day the difference in norovirus concentrations estimated from the CrAsspahge concentrations would be approximately 10% higher if differential decay was considered. Futher data is required to examine the decay rates in freshwater, seasonal differences, and the impact on norovirus infection or illness estimates.

The above simulation model and example results show CrAssphage is a promising indicator for identifying differences in norovirus infection rates. Further analysis is required to improve understanding the uncertainty associated with these estimates and the effect of decay rates.

14.3 *E. COLI* AS A SCREENING TOOL FOR CRASSPHAGE

The survey data for water samples with greater than 100 GC CrAssphage/100 ml is plotted against *E. coli* in Figure 36. The *E. coli* concentration does show a relationship with the observed range of CrAssphage concentrations. Increasing *E. coli* corresponding to an increasing upper bound of observations of CrAssphage concentrations.

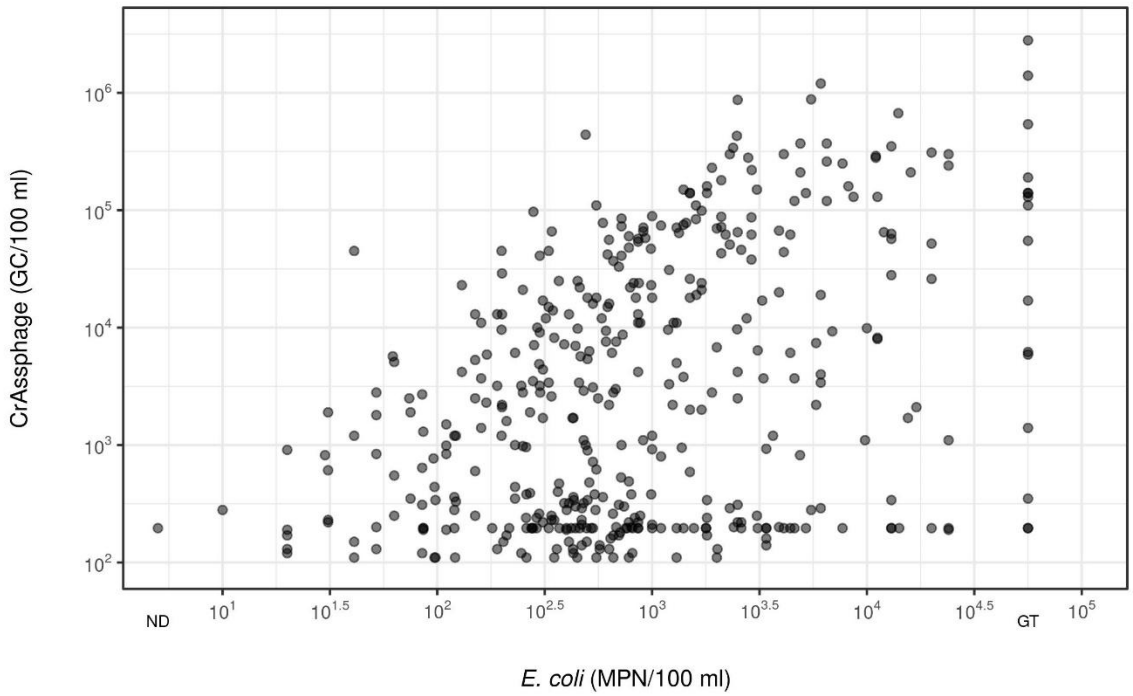


Figure 36 Scatterplot of CrAssphage concentration (GC/100ml) compared with *E. coli* (MPN/100ml)

Samples with elevated CrAssphage, but low *E. coli* were all from rivers which had at least 33% and up to 96% of the other samples from each river with high levels of *E. coli.* Many samples with high *E. coli* concentrations will not have high CrAssphage due to the *E. coli* being from other sources such as wild bird or ruminants. While this means *E. coli* has a low specificity for CrAssphage, it has a very high sensitivity - ie most samples with CrAssphage which exceed RBT, will exceed the guideline *E. coli* thresholds.

Enterococci was less useful for predicting the presence of CrAssphage, particularly in the region of interest of 10^4 to 10^5 GC CrAssphage/100 ml (Data not shown).

to

PART D

Discussion and Conclusions

15 COMPARISONS TO THE 2002 FRESHWATER PROGRAMME

This study provides an improved and expanded data set describing the relationships in New Zealand rivers between a range of microbial indicators and pathogens. This section compares both the methodological approaches for the survey, the QMRA simulation model approaches, and the results obtained.

15.1 DATA COLLECTION AND LABORATORY ANALYSIS

Table 52 provides an overview of differences in laboratory analysis and results obtained in the river surveys undertaken in 1998-2000 (FMRP) and 2020-2024 (current study).

The current study collected 44% more samples, from a larger number of sites. This included proportionally twice as many urban samples, collected from 13 sites, rather than just three, providing a wider representation of urban contamination. MST analysis (which was not available for the FMRP study), confirmed human MST markers in 35% of samples. The FMRP study included more forestry/undeveloped sites (28% samples), compared with the current study low impact sites (9% samples).

This site selection difference may explain why for the FMRP study, 50% of samples had less than 100 MPN *E. coli*/100 ml, compared with the current study where the median was 310 MPN *E. coli*/100 ml. This provides more data in the areas of highest interest but does also mean that the samples were on average more contaminated in the current study. This doesn't mean that New Zealand freshwater sites have become more contaminated, rather the freshwater sites were selected to include those more likely to be contaminated.

For every organism except for *E. coli*, there were methodological differences in the laboratory analysis undertaken between the two studies. For the pathogens, methodology used in the current study is almost certainly better than that used in the FMRP. Over the last 25 years there has been a considerable increase in understanding and experience in isolation of pathogens, and in the use of molecular tools including qPCR and sequencing of DNA.

The FMRP study only determined quantitative levels of thermotolerant *Campylobacter*, which when adjusted for differences in detection limits were similar to the current study (60% vs 65%). There were similar levels of thermotolerant *Campylobacter* except between 10 to <100 MPN/100 ml where the current study had a higher proportion of samples, and with samples > 100 MPN / 100 ml, where the FMRP had a higher proportion of samples.

In the FMRP speciation of the *Campylobacter* was undertaken at a later point on frozen extracts. In the FMRP, in just under half of the samples with *Campylobacter* these were identified as *C. jejuni,* whereas in this study, in 94% of the samples with *Campylobacter* these were identified as *C. jejuni*. The FMRP study noted that a number of samples reanalysed were likely to have degraded during storage, as reanalysis identified no *Campylobacter*. Therefore, more samples may have actually had *C. jejuni*. The greater proportion of forestry/undeveloped samples in the FMRP may also have created bias towards other species of *Campylobacter* which might be more common from wild bird and wildlife.

Salmonella were isolated from a higher proportion of samples in this study than in the FMRP study (16% vs 10%). Noting the association with human MST markers, the greater proportion of samples with human inputs may contribute to this. *S.* Brandenburg was the most frequently detected serotype in FMRP study, which at the time was attributed to an outbreak at the time involving sheep, birds and human cases (Baker et al, 2007). The 11 samples from FMRP with > 1 MPN/100 ml were all from sheep or bird sites. While the current study only enumerated *Salmonella* by culture in the pilot study, when combined with qPCR, the current study suggests consistency with FMRP results in that most detections of *Salmonella* are at less than 1/100 ml.

Cryptosporidium oocysts and *Giardia* cysts were identified in a greater percentage of samples in the current study, and at higher concentrations (Table 46).

The 1998-2000 study used a method that was promulgated by USEPA in 1996 - the Information Collection Rule (ICR) protozoan method, which used nominal porosity filtration and indirect fluorescent antibody staining (USEPA., 1996). While it was the best available method at the time, it was subsequently deemed by a number of researchers to be problematic with regard to having both low and variable recoveries following the analytical process (Clancy et al, 1994; Connell et al, 2000; Hsu and Huang, 2000).

In order to try to increase test performance, Massey Protozoa Research Unit amended one of the processing steps by replacing the sucrose flotation clean-up with immunomagnetic separation (IMS) which is now a mandatory step. However, other sources of variability were still evident with the ICR method (Jennifer L. Clancy et al, 1999).

USEPA method 1623 (and its predecessor method 1622), which was used in the current study, has been extensively trialled and validated to highlight improved recoveries compared to ICR method (J. L. Clancy et al, 1999; DiGiorgio et al, 2002; Ferguson et al, 2004; McCuin and Clancy, 2003). One study showed that the protozoan recovery efficiencies of Method 1623 are about 5 times higher than the ICR method (Hsu et al, 2002).

In the 1998-2000 FMRP study enterovirus were detected in 33% of samples and adenovirus in 32%. Only 11% of samples had both enterovirus and adenovirus detected in the same sample, so overall 54% of samples had enterovirus and/or adenovirus. This analysis was presence/absence only with no quantitation possible. These detections were across all sample types (bird, dairy, forestry, sheep and municipal) with municipal sites ranking only fourth.

In this study, pathogenic viruses were only tested for in phase I, in 52 samples. Despite the small sample size, the contrast with the FMRP study is quite stark, with no detections of adenovirus, and only 2 samples (4%) with very low-level detections of enterovirus. The explanation for the much lower level of detections is not obvious. Laboratory analysis during the pilot study in 2020 is likely to be more sensitive than in the 1998-2000 study, not less. The sites in the pilot study were also selected to represent higher levels of contamination than the FMRP study, with MST markers confirming human sources of faecal pollution in almost half of the samples tested in the pilot. In the current study, the MST marker CrAssphage (which was not discovered until 2014) provides useful quantitative data for the potential presence of pathogenic viruses, in lieu of actual viruses.

Overall, the current study provides a larger and more robust data set, with many more samples with elevated concentrations of faecal indicators and pathogens. This richer data set supports more rigorous QMRA modelling.

Table 51: Comparison of key sampling and laboratory analytical differences between the 2002 and 2025 studies.

Component	2002	2025
Number of sites	25 sites each sampled 29 times	72 sites each sampled 1 to 38 times
Number of samples	725	1,041
Duration of sampling	15 months (Dec 1998 to Feb 2000)	30 months (Feb 2020 – Jun 2024)
Categorisation of sources impacting water quality	Observed land use or activity	Observed land use or activity and MST markers
Observed land use or activity Number of sites	Municipal (M) (+D, F and S) (3 sites) Dairy (D) (+F and S) (5 sites) Sheep (S) (+D and F) (6 sites) Birds (B) (+M, F and S) (4 sites) Forestry/Undeveloped (F) (7 sites)	Urban (U) (13 sites) Dairy (D) (22 sites) Sheep and Beef (SB) (16 sites) D and SB (2 sites) Wild bird (W) (+U, D and SB) (10 sites) Low Impact (L) (8 sites)
Number of samples	Municipal (M) (+D, F and S) (87) Dairy (D) (+F and S) (145) Sheep (S) (+D and F) (174) Birds (B) (+M, F and S) (116) Forestry/Undeveloped (F) (203)	Urban (U) (261) Dairy (D) (236) Sheep and Beef (SB) (221) D and SB (53) Wild bird (W) (+U, D and SB) (176) Low Impact (L) (94)
MST Medium or higher levels of marker	Not available	Human MST markers (35%) Ruminant MST markers (66%) Wild bird MST markers (80%)
Faecal Indicator Bacteria		
Total coliforms	Not Reported	Colilert (10 ml) <10 to 24,000 MPN/100 ml
E. coli - Methodology Volume tested Detection range	Colilert 100 ml <1 to 2,400 MPN/100 ml	Colilert 10 ml <10 to 24,000 MPN/100 ml
E. coli - Results	Detected 99% samples. 9 sample <1 /100ml Median 100 MPN / 100 ml 24 samples >2,400/100 ml (samples above this concentration were not enumerated)	Detected 99% samples. 15 samples <10 /100ml Median 310 MPN / 100 ml 123 samples > 2,400/100ml, which due to 1:10 dilution all but 17 (with> 24,000/100ml) were enumerated
Enterococci - Methodology Volume tested Detection range	Not tested	Enterolert 10 ml <10 to 24,000 MPN/100 ml
Clostridium perfringens spores | Somatic and FRNA coliphage	Membrane filtration, TCS-fluorocult agar | Single layer plaque assay on WG5 and WG49	Not tested
Campylobacter		
Thermotolerant *Campylobacter* -	2 stage 3x3x3 MPN in Prestons Broth with detection by PCR of	1x3x3x3 MPN enrichment in Bolton Broth with broth plated onto modified

Component	2002	2025
Methodology	thermotolerant species using Eyers et al (1993) thermophilic 23S primers (290 bp amplicon).	Charcoal-Cefoperazone-Deoxycholate Agar (CCDA) plates. Where colonies grew, broths were screened using multiplex PCR (Wong et al. 2004) with thermophilic 23S (246 bp amplicon).
Detection Range	<0.3 to >110 MPN/100 ml	<0.1 to >1,100 MPN/100 ml
Thermotolerant *Campylobacter* MPN/100 ml	 Detected 59.5% (DL=0.3/100 ml) 23% < 1 MPN (164/725) 20% 1 to <10 MPN (146/725) 8% 10 to <100 MPN (57/725) 9% > 100 MPN (64/725)	Detected 68% (DL=0.1/100 ml) Detected 65% (DL=0.3/100 ml) 22% < 1 MPN (226/1028) 21% 1 to <10 MPN (211/1028) 20% 10 to <100 MPN (206/1028) 5% > 100 MPN (52/1028)
Campylobacter species identification - Methodology	Subsequent PCR retest of the largest positive MPN broth to identify *C. jejuni, C. coli* and *C. lari*.	Every sample tested with multiplex PCR that included primers also for *C. jejuni and C. coli.* Whole genome sequencing of isolates
Proportion of samples with *Campylobacter* which were identified as	*C. jejuni* - 48% *C. coli* - 1% *C. lari* - 19% Others – 22% Unknown - 17% (These presence/absence without quantitation)	*C. jejuni* - 94% *C. coli* - 20% *C. lari* - 1% Others – 3%
Salmonellae		
Salmonellae - Methodology	MPN (830 ml) resuscitation in peptone enrichment broth, followed by enrichment in Rappaport-Vassiliadis Soya (RSV) and Selenite Cysteine broths, with TECRA LISA positive broths plated on Xylose Lysine Deoxycholate (XLD) agar. Confirm by ELISA or biochemical properties Detection limit (0.12/100 ml)	MPN (903 ml) resuscitation in peptone enrichment broth, followed by enrichments in Muller-Kauffmann Tetrathionate Novobiocin, and RSV broths, before plating to XLD and Hektoen Enteric Agar. Confirmed by biochemical properties Detection limit (0.11/100 ml). Phase 2 and 3 single 1 L sample analysed.
Detection range	0.12 to 11 / 100 ml	0.11 to 110 / 100 ml
Salmonellae detections	10% of samples 5 samples > 10 MPN/100 ml 6 samples 1-10 MPN/100 ml 55 samples < 1 MPN/ 100 ml	16% of samples Maximum 0.25 MPN/ 100 ml
Salmonellae serotypes	*S.* brandenburg (30%) *S.* typhimurium (14%) *S.* hindmarsh (4%) *S.* Mississippi (1%) Others (19%)	*S.* brandenburg (2%) *S.* typhimurium (57%) *S.* hindmarsh (4%) *S.* Mississippi (1%) *S.* Bovismorbificans (12%) *S.* Enteritidis (10%) *S.* Give (7%) Others (30%)

Component	2002	2025
	Unable to be typed (32%)	
STEC	Not measured	Method provide in Appendix
Viruses		
Viruses - Methodology	10 litre sample concentrated using Virasorb filters, with 1 litre equivalent analysed by PCR for presence/absence	10 litre sample concentrated using hollow-fibre ultrafiltration, with beef extract elution, followed by PEG precipitation, and extraction using Presto Extraction kit and RT-PCR.
 Adenovirus Enterovirus Norovirus GI Norovirus GII	 Detected 32% samples. Detected 33% samples. Not tested Not tested	Phase one only (52 samples) Detected 0% samples. Detected 0-4% samples. Detected 4-12% samples. Detected 6-13% samples.
Protozoa		
Protozoa -Methodology	Wound polypropylene thread cartridge filtration (CUNO MCIROWYND) of 100L and immunomagnetic bead recovery and visualisation by microscopy.	IDEXX Filta-Max® filtration followed by immunomagnetic bead recovery and visualisation by microscopy.
Cryptosporidium oocysts Per 10 L	Detected 5% samples. 0.1 to 0.9 oocysts / 10 L – 29 samples ≥1 oocysts / 10 L – 4 samples Maximum – 11.9 oocysts / 10 L	Detected 50% samples. 0.1 to 0.9 oocysts / 10 L – 158 samples ≥1 oocysts / 10 L – 334 samples Maximum – 140 oocysts / 10 L
Giardia cysts Per 10 L	Detected 8% samples. 0.1 to 0.8 cysts/ 10 L – 49 samples ≥1 cysts/ 10 L – 10 samples Maximum – 8.4 cysts/ 10 L	Detected 69% samples. 0.1 to 0.8 cysts/ 10 L – 146 samples ≥1 cysts/ 10 L – 535 samples Maximum – 94 cysts/ 10 L
	18% of samples with *Cryptosporidium* also had *Giardia* detected. 10% of samples with *Giardia* also had *Cryptosporidium* detected	76% of samples with *Cryptosporidium* also had *Giardia* detected. 55% of samples with *Giardia* also had *Cryptosporidium* detected.

15.2 COMPARISON OF THE 2002 AND 2025 CAMPYLOBACTERIOSIS QMRA MODELLING APPROACHES

15.2.1 Model components

This section compares the QMRA model developed in 2002 using freshwater field data generated in New Zealand between 1998 to 2000 (McBride et al., 2002) and the model described in sections 10 to 13. The model components of the two studies are summarised in Table 52 and Figure 37.

Both studies use Monte Carlo simulation which remains a useful and established tool for this type of quantitative risk assessment. Since 2000, improved data for water intake during swimming activities has been published, which are implemented in the 2025 model.

A key difference between the two models is the information flow through the models as shown in Figure 37. In the 2002 model, the risk of *Campylobacter* infection is based on the complete survey data for *Campylobacter*. After the infection outputs are calculated, the infection outputs are matched to *E. coli* concentrations via percentile matching.

In the 2025 model, quantile regression allows the highly variable relationship between *Campylobacter* and indicators to be captured directly. This relationship can then be used to define a distribution of the expected *Campylobacter* concentrations that could be present in water given a known *E. coli* concentration. The improved modelling of the relationship between faecal indicators and *Campylobacter* concentrations is possible due to the advances in computing power since 2002.

The 2002 study used infection as the reported metric, whereas in 2025 infection, illness and DALY risk metrics have been determined. In the absence of new infection or illness data for low doses of *Campylobacter,* the dose-response model for infection used in 2002 is also used for the 2025 model.

Infection estimates in 2002 were determined using (a) Percentage of swims resulting in infection for 1000 people at 1000 different beaches a day; and (b) Percentage of swims resulting in infection for 1000 people at the same beach each day. Infection estimates were made using all the data or just bathing season data.

In 2002, approach (a) provided insufficient discrimination, so the second metric (1000 people swimming at the same beach each day) was used to inform the Guidelines.

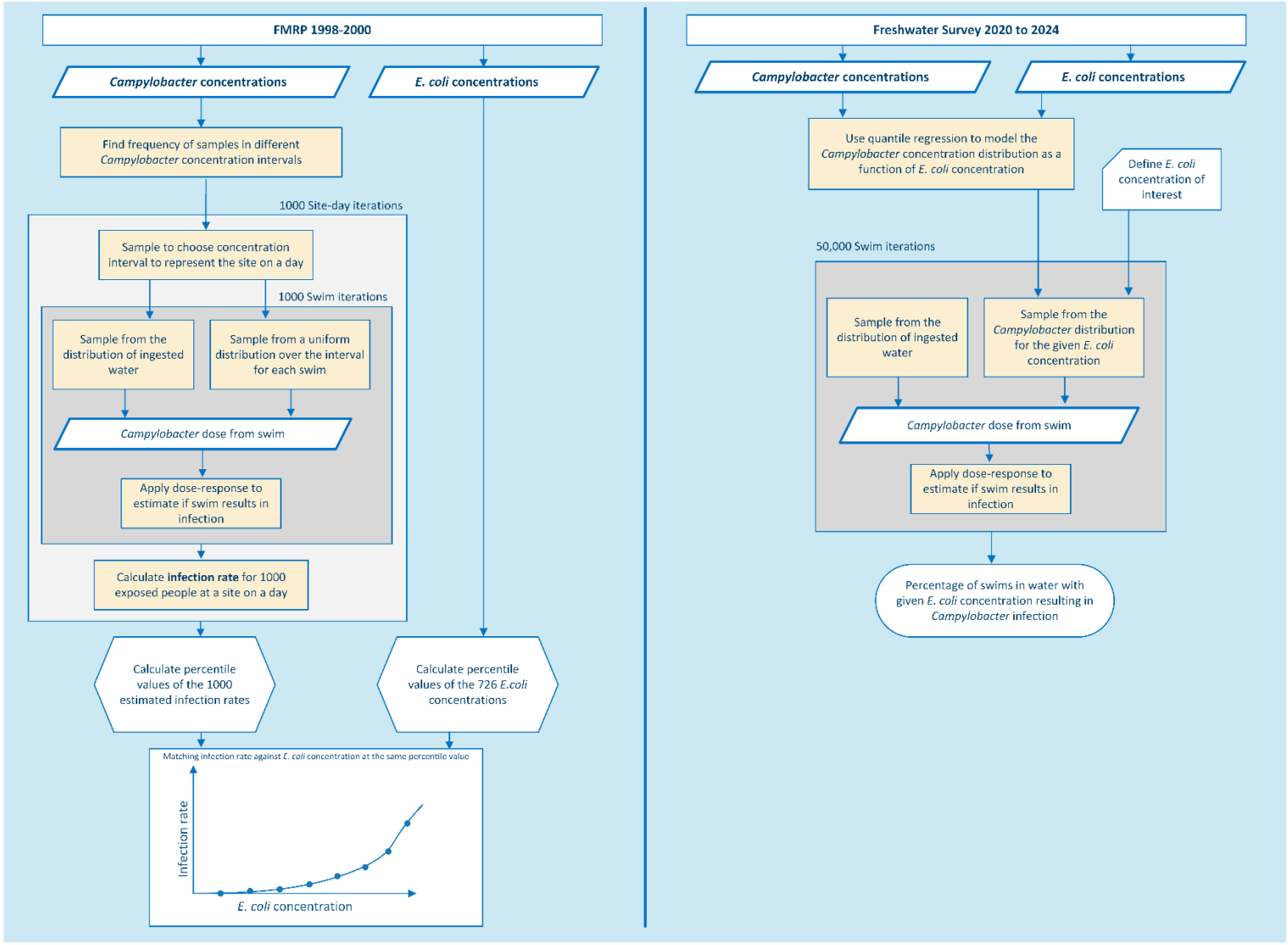
In 2025, with the larger dataset the modelling approach used is more similar to 1000 people at 1000 different beaches.

Both of these models are designed as screening tools, where *E. coli* is being used to categorise the water and can be used across the country in the absence of extra supporting data to categorise the risk. The FMRP did not analyse samples for enterococci. The inclusion of this indicator in the current study allowed the models to be also run using enterococci.

Table 52: High level comparison of the QMRA simulation models reported in 2002 and 2025.

	2002 Model	2025 Model
Pathogen	Thermotolerant *Campylobacter*	*Campylobacter jejuni* or *coli*
Indicator bacteria	*E. coli*	*E. coli* and enterococci
Focus of water sample collection	Nationally representative samples, covering sites with Municipal, Dairy, Sheep, Bird and Forestry/undeveloped observed land types.	Nationally representative survey of geographically spread sites that covered a range of climatic conditions, faecal sources, and had adjacent land activities likely to impact freshwater quality.
Simulation model type	Monte Carlo simulation model	
Representing *Campylobacter* concentrations	Sampled from uniform distributions for defined intervals of *Campylobacter* concentration. The frequency of choosing each interval is based on the relative frequencies of the *Campylobacter* concentrations in the survey. *Campylobacter* concentration is independent of *E. coli* concentration.	Uncertainty in MPN estimates, represented by lognormal distribution of *Campylobacter* concentration. *Campylobacter* distributions based on quantile regression of the *Campylobacter* concentration (incorporating uncertainty) as a function of the *E. coli* concentration. Regression analysis using the 2020-2024 survey data.
Each simulation iteration represents	One thousand swims at a site on a day with a defined single interval of *Campylobacter* concentration. OR One thousand swims at 1,000 different sites per day.	A swim at a site with a *Campylobacter* concentration that could be expected for a defined indicator concentration.
Number of simulation iterations	1,000 iterations	50,000 iterations
Description of risk from campylobacteriosis	Percentiles of risk of *Campylobacter* infection from all data in survey.	Percentage of swims resulting in *Campylobacter* infection or illness, while swimming in water with a given *E. coli* concentration. Plus DALY per 1,000 swims.
Understanding the relationship between the indicator and the pathogen	Percentile matching of the infection rates with *E. coli* concentrations from the survey.	The relationship between the indicator and *Campylobacter* is defined before running the simulation model. Results are given for a defined indicator value.
Water ingestion	Distributions estimated from swimmers and primary contact water activities	Updated distribution based on 6 to 12 year old open-water swimming.
Dose-response	Medema *et al.* (1996) model	

Figure 37: Flow diagram showing the structure of the 2002 and 2025 simulation mode



15.2.2 Model results

The results from 2002 model (1,000 people swimming at the same site each day and using all data) were used to inform the 2003 Ministry for the Environment and Ministry of Health Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas (the Guidelines). The results for the 2002 and 2025 models at the three *E. coli* concentrations used in in the National Policy Statement for Freshwater Management 2020, are shown in Table 53.

Table 53: Comparison of the percentage of swims resulting in *Campylobacter* infection (2002 model) and infection or illness (2025 model) for swimming in water with given values of *E. coli* concentrations.

Model	2002 FMRP				2025		2025	
	Percentage of swims resulting in infection for 1000 people				Percentage of swims resulting in infection		Percentage of swims resulting in illness	
	at 1000 different beaches a day		at the same beach each day					
E. coli / 100ml	All data	Bathing season	All data	Bathing season	All data	Bathing season	All data	Bathing season
130	4.2%	3.3%	0%	0%	3.5%	2.0%	1.5%	0.5%
260	4.4%	3.5%	0.9%	0.4%	5.5%	4.0%	2.0%	1.5%
540	4.6%	3.7%	4.9%	2.5%	8.5%	7.0%	3.5%	3.0%

In the 2002 model, which used percentile matching rather than the direct relationship between *E. coli* and *Campylobacter*. The percentile matching after calculating the infection rate makes it possible to estimate a zero risk. Forty percent of the iterations (1,000 swimmers at one site each day) would have corresponded to swimming in water with between 0 and 1 *Campylobacter*/100 ml, concentrations unlikely to cause infection. Sixty three percent of iterations corresponded to less than four *Campylobacter* / 100 ml. For context the 65th percentile of the 2002 survey *E. coli* concentrations was 191 *E. coli* / 100 ml.

The alternate approach used in the 2020 to 2024 survey, was to predominately sample from sites with an expectation to sample faecal contaminated water. This approach minimised the number of samples where no pathogens were detected and provided a good spread of indicator and *Campylobacter* concentrations. *Campylobacter* concentrations that could cause infection were observed even at low *E. coli* levels, so the 2025 model does not estimate a zero risk of *Campylobacter* infection for low *E. coli* levels.

The 2025 risk estimates are not over inflated because of the decision to focus sampling on known contaminated sites. The model uses the relationships between the indicator and *Campylobacter*, rather than the number of sites with different concentrations in estimating infection rates.

16 DISCUSSION AND CONCLUSIONS

Recreational water guidelines are a policy decision based on societal views and consideration of the actual and potential health risks from contaminated water. This report updates the understanding of the science underpinning the assessments of the health risks from microbial contamination of New Zealand rivers.

Consistent with expert opinion 25 years ago, for New Zealand, a QMRA remains the most practical and useful mechanism of assessing human health risk posed by microorganisms in rivers (McBride et al. 2002). Also, as was determined 25 years ago, it was necessary to generate New Zealand-based data describing the pathogens and indicators in our rivers. The data set of over 1,000 water samples with measured microbial and environmental parameters provides a sound foundation for discussions over appropriate microbial water quality management in New Zealand.

16.1 CAMPYLOBACTER

Campylobacter was the most commonly isolated bacterial pathogen in this study, being isolated from 68% of samples, with 94% of these *Campylobacter* being *C. jejuni.* The *C. jejuni* were also present at concentrations of significance to human health, with 4% of samples with greater than 100 *C. jejuni* per 100 ml, 19% with between 10 and 100 *C. jejuni* per 100 ml, and 19% between 1 and 10 *C. jejuni* per 100 ml.

There are relatively few studies which can be directly compared with the current study. Overseas studies such as those in Australia (Henry et al, 2015), Canada (Guy et al, 2018), Japan (Kobayashi et al, 2022), and South Africa (Chibwe et al, 2023) have reported prevalence in surface water ranging from 8 to 90%. In the studies where *Campylobacter* were enumerated, concentrations up to 100 MPN/100 ml have been observed, but similar to this study, most concentrations were much lower. A number of these studies were based on a single location with small sample numbers. Overseas conditions and sources of contamination can be very different to New Zealand, so previous New Zealand studies are more relevant to this discussion.

Since the 1998-2000 survey, while there have been New Zealand studies published, none have been quantitative, instead using presence/absence methodology. There have been a range of observations from 20% of samples positive for *Campylobacter* (Phiri et al, 2020), through to 55% of samples from rivers in the Ashburton region (Devane et al, 2005), and up to 90% of samples from three rivers in the Manawatu (Cookson et al, 2024).

None of these studies were as comprehensive as the current study, but they do illustrate a significant range of *Campylobacter* prevalence, and as in this study, the prevalence will differ between rivers according to faecal contamination sources. Where determined, the concentrations of *Campylobacter* in previous studies were within the range observed in this study.

C. jejuni is the most frequently isolated species in published studies, although in some studies other species can be more common than observed in this study. Cookson et al. (2024) found 39% of isolates were *C. coli*, compared to only 17% in this study.

The relatively low proportion of *C. jejuni* and *C. coli* reported in the FMRP study (49% of *Campylobacter*) could have suggested an overestimation of risk when the QMRA was based on concentrations of total *Campylobacter* spp. Additional species of *Campylobacter* are increasingly being identified (Carter et al, 2009; Cookson et al, 2024; Costa and Iraola, 2019), some of which were found in the current study. There is little evidence for these other *Campylobacter* species being a significant health risk (Costa and Iraola, 2019). The high proportion of *C. jejuni* in the current survey (94%), supports that it is likely that the 1998-2000 study underestimated the proportion of *C. jejuni* in the *Campylobacter*-positive samples. This could have been influenced by the species identification step being subsequent to the initial analysis where degradation of the samples during storage resulted in no species being identified, rather than a *Campylobacter* species being detected. The high proportion of *C. jejuni* in the current study supports the pathogenic potential of the *Campylobacter* found in rivers in both the 2002 and 2025 freshwater projects. Land use changes, and changes in the carriage of *Campylobacter* will also have contributed to changes in the relative prevalence of *Campylobacter* species.

In this survey, *C. jejuni* were associated with human, ruminant and wild bird sources, with higher prevalence and higher levels associated with samples with multiple faecal sources rather than a single source. Ruminants and poultry are well documented as being sources of *C. jejuni* (Devane et al, 2005; Gilpin et al, 2008; Rapp et al, 2014). The high incidence in humans also means that human sewage will normally contain *C. jejuni*, although concentrations will fluctuate with disease incidence in a community.

Whole genome sequencing was used to further characterise the *Campylobacter*. Almost half of the *C. jejuni* have a sequence type that either has been found in human clinical cases, or from poultry or ruminants which are known zoonotic sources of infection. We observed the highest proportion of these previously documented *Campylobacter* STs from samples with a human only MST profile (Table 42). Conversely, the greatest proportion of novel STs were observed in samples with no source identified suggestive of wildlife faecal sources such as possums. While qualitative estimates of infectivity of *Campylobacter* from non-human sources have been made (Soller et al, 2010), as noted in a recent USEPA report (USEPA, 2024), there is insufficient data available to assign an infectious fraction of *Campylobacter* from non-human sources. Should this data gap be addressed, then reanalysis of the data in this study could be undertaken.

Campylobacter in poultry, ruminants and wild birds are primarily commensal organisms which usually cause few illness symptoms (El-Saadony et al, 2023). For the human population, where *Campylobacter* infection frequently causes disease (Igwaran and Okoh, 2019), poultry are the primary source (Lake et al, 2021). Commercial poultry at birth are usually *Campylobacter* free, but when sheds become infected, the whole flock is often infected with the same sequence types. When the human population is infected from a food source that is widely distributed, then that ST will be common in the human population. ST45, which has been isolated from poultry and human clinical cases (Lake et al, 2021), was the second most frequently isolated ST (62 isolates from 30 sites) from the water samples. There were clusters of similar genotypes from across the country (Fig 10).

Many of the other genotypes, including ST-2381, which was the most common sequence type in this study, have only been isolated from New Zealand rivers, and/or wild birds (Carter et al, 2009; Shrestha et al, 2019). ST-2381 was the most common ST identified in river

samples from the Ashburton region sampled in 2000-2001, and from rivers in the Manawatu region sampled in 2008 (Carter et al, 2009; Shrestha et al, 2019).

Whether ST-2381 and other ST derived from either wild birds, or an unknown source are any less pathogenic to humans is unknown because very few *Campylobacter* in New Zealand isolated from clinical cases have sequence types determined. What data we do have on clinical cases is primarily from research studies undertaken over defined periods in Auckland (Lake et al, 2021), Christchurch (Gilpin et al, 2013) and Palmerston North (Müllner et al, 2010), and from outbreak investigations (Gilpin et al, 2020). While sequencing of a representative number of clinical isolates could resolve this issue, in the absence of this, it is prudent to assume that the majority of *C. jejuni* and *C. coli* identified in rivers are able to cause disease in humans.

16.2 OTHER PATHOGENS

Cryptosporidium and *Giardia* were detected in 50% and 69% of samples respectively, but only 1% of samples had greater than 10 *Cryptosporidium* per 10 L and 3% of samples had more than 10 *Giardia* per 10 L.

Very few STEC were isolated by culture. STEC genes were detected in 307 samples (31%), but only 13 isolates were sequenced that were STEC. Some of this reflects the low selectivity of culture methods meaning that many other bacteria present on isolation plates mask the STEC. This is compounded by the low levels of STEC present. The maximum concentration in the initial MPN enrichment was 0.25 MPN per 100 ml, while qPCR quantitation in all but one sample was less than 930 GC per 100 ml. STEC genes in broths may also be present in other species of bacteria, suggesting non-target amplification. For future work, enrichment steps seem necessary, either with qPCR or sequencing of broths, or use of more selective media for isolation.

Salmonella were detected in up to 19% of samples, but concentrations were almost always <1 per 100 ml.

QMRA analysis was not feasible for STEC and *Salmonella*.

16.3 MICROBIAL SOURCE TRACKING

Microbial source tracking has emerged as a key technology for understanding water quality and tracking sources of faecal contamination (Devane et al, 2018). There are at least three providers of microbial source tracking services in New Zealand (Watercare, Cawthron and ESR), and widespread use of these services by councils for source identification.

This study provided an opportunity to use MST to better characterise the sources of contamination in a river beyond what was visible from site inspections and sanitary surveys. This study also provided the opportunity to evaluate MST markers alongside indicators and pathogens.

Microbial source tracking markers for human faecal material were identified in 35% of samples, ruminant faecal sources in 66% of samples, and wild bird faecal markers in 80% of samples. Seventy percent of samples had more than one faecal source. The use of these tools highlighted the intermittent nature of faecal pollution at many sites, and the contribution

of unexpected sources including non-livestock animals identified through environmental DNA analyses.

MST tools provide a second level of analysis, that is best applied when elevated levels of faecal indicators prompt further investigation. In the case of human sources being identified, the concentrations of the human MST marker, CrAssphage, can be used to estimate the additional risk from viruses such as norovirus. Only the first 52 samples in the study were tested for pathogenic viruses, and only very low levels were detected, despite many of the samples being identified as contaminated with human sewage. The low recovery of pathogenic viruses from environmental matrices and their low concentration contribute to this lack of detection, however, these factors do not mitigate their health risk potential. The likelihood that this viral analysis would not generate data that could be used in QMRA, together with the cost of analysis, and commitments of the testing laboratory to wastewater testing for SARS-CoV-2, resulted in a decision to discontinue testing for pathogenic viruses. However, as part of the MST analysis, the human faecal indicator CrAssphage was evaluated to assess viral risk from human sources. This emerged as a viable approach during this study with publications by researchers from USA and Australia (Ahmed, Schoen, et al, 2024; Boehm and Soller, 2020; Crank et al, 2019; Schoen et al, 2020). There has also been considerable analysis of human sewage recently (due to sampling related to the Covid pandemic) which has generated significant data supporting the relationship between CrAssphage and pathogenic viruses including norovirus. We adapted the most recent QMRA for CrAssphage to estimate risk of illness from norovirus (Ahmed, Schoen, et al, 2024).

The positive association of *E. coli* with CrAssphage means that having identified samples with elevated *E. coli* levels, MST analysis can then be undertaken. When human sewage is identified, then CrAssphage concentrations can be used to estimate health risk. Human faecal pollution of rivers has additional risks of infection from human infective viruses. QMRA analysis based on the human MST marker CrAssphage suggests illness risks from norovirus of >3% when 30,000 genome copies/100 ml of CrAssphage are detected. This risk is in addition to the risk identified in the *Campylobacter* based QMRA and emphasises the importance of human sewage as a risk to human health.

While a new technology, the use of eDNA analysis has become widespread across New Zealand with the primary focus on understanding biodiversity in rivers, and both invasive and native species. In this study eDNA analysis indicated contamination from 19 different mammals (cattle, rat, possum, sheep, and deer the most frequently observed), and 40 different wild birds (mallard duck and pūkeko the most frequently observed). Metagenomic analysis such as eDNA is not quantitative. The persistence of DNA from animal or bird cells may also mean that detections are not recent. Detections may also reflect the diet of animals or wild birds in or around a river. The eDNA does highlight the need to consider a wide range of potential sources of faecal contamination, and that eDNA analysis undertaken for other purposes, can be leveraged to assist water quality management.

16.4 QMRA APPROACH TO SUPPORT NATIONAL LEVEL METRICS

The QMRA as detailed in sections 10 to 13 is part of the toolbox that can be used for setting national guideline values for recreational water sites. In general, the main strengths of a QMRA are the transparency of the approach, it contains the key exposure and response components and the QMRA model provides quantitative estimates of risks to human health

(WHO, 2016, Bichai and Smeets, 2013). The process of developing a QMRA has the advantage of identifying knowledge gaps and supports understanding of how the choice of components of the model affect the risk estimates (Ashbolt et al, 2010).

The quality of a freshwater recreation QMRA depends on the quality and availability of data on:

- The behaviour of people recreating at the freshwater site, including the frequency of exposures and how much water is ingested.
- How pathogens and indicators survive or are diluted in the water following entry into the water.
- The probability pathogens ingested by a person resulting in a disease infection or illness.

The study QMRA analysis used Monte Carlo simulation models allowing the use of variability distributions to describe data, replicating the variability seen in the real world. The simulation model estimates the infection and illness risk for swimmers at freshwater sites, accounting for the variability of swimmer behaviour and concentrations of indicators and pathogens in the water. Simulation models also allow incorporation of uncertainty in model parameters, as used for some dose-response relationships.

Where there is uncertainty whether the chosen model data reflects the New Zealand situation, sensitivity analysis has been conducted to determine how the choice of data impacts the estimates of disease outcomes (section 13, appendix E and F).

16.4.1 Single verses multiple pathogen QMRA estimates of public health risk

Multiple pathogen risk assessments are based on either epidemiological studies with a gastrointestinal or respiratory illness endpoint that does not identify specific diseases (Kay et al, 1994; Wiedenmann et al, 2006) or use QMRA to combine the illness risks from multiple pathogens (Boehm and Soller, 2020) or consider the risk from each pathogen separately (Schijven et al, 2011).

A New Zealand epidemiological study to understand the general risk from pathogens from swimming at recreational water sites would need to involve a very large number of study participants in order to provide statistically robust predictions of relationships between indicators and illness (USEPA, 2024, Soller et al, 2016). The large sample size would be driven by the low rates of illness estimated by this study's simulation model, the typically low pathogen concentrations measured in water samples, and the low numbers of people swimming at individual freshwater sites in New Zealand compared to international locations. For these reasons a New Zealand study would require a multi-year study at multiple sites, and would not be logistically or financially feasible.

Overseas epidemiological studies have typically been focussed on human impacted sites. In comparison, this study has shown freshwater in New Zealand can frequently be impacted by multiple sources, so any relationships between indicators and illness developed internationally may not be applicable to New Zealand freshwater.

It is possible to use this study data and models to provide combined illness estimates for campylobacteriosis, giardiasis and cryptosporidiosis for the purpose of developing national guidelines. However, for the following reasons the model calculates infection and illness rates for single rather than multiple pathogens:

- Concentration data is required to determine risk profiles. This study collected data for only three of a large cohort of possible pathogens that could be in fresh water. The combined probability of illness from *Campylobacter*, *Cryptosporidium* and *Giardia* would not account for other possible pathogens.
- The illness rates from *Campylobacter* are greater than for *Cryptosporidium* and *Giardia*.
- The uncertainty in disease outcome estimates would be increased when combining the uncertainties from the individual dose responses.

16.4.2 Water ingestion

The volume of water ingested by people at freshwater recreational sites on a visit is highly variable and depends on the activities undertaken. In the QMRA, the volume ingested is derived from distributions for the ingestion rate per hour and the duration of the activity. Swimming is chosen as an activity that is likely to cause the greatest unintentional ingestion of water, and this is line with other international QMRA studies (Ahmed et al, 2024, Skiendzielewski et al, 2024 and Soller et al, 2014).

Sensitivity of the health outcome rates to the choice of water ingestion approach is described in appendix F.3. The study ingestion distribution provides similar rates of *Campylobacter* infection and illness, using a fixed 40-50 ml ingested volume.

There are no New Zealand specific data for the duration of activities or ingestion rates during swimming at freshwater sites. The study QMRA uses overseas data (Dufour et al, 2017, DeFlorio-Barker et al, 2018) to estimate ingestion during swimming, and uses children aged 4-12 year old as the population of interest as a conservative approach for supporting national analysis. The 4-12 year age-group ingests the most water and spends more time in the water than other age groups. For example, in the DeFlorio-Barker et al. study (2018) 4-12 year old children spent a median of 60 minutes (mean of ~90 minutes) in the water compared to adults aged 35 years and over who spent a median of 30 minutes (mean of ~50 minutes) in the water.

Epidemiological studies in the United States confirm children are at a higher risk of illness than adults from swimming at recreational water sites (Wade, 2022). However, the relative effects on illness outcome from differential dose response and water ingestion is unclear.

Members of the Governance and Stakeholder groups questioned whether the water ingestion volumes may be different for different parts of the country or for different population groups. Whether people in different parts of the country spend more or less time swimming is unknown. Warmer parts of the country may have longer “swimming” seasons, but there is a growing trend towards ice swimming or wild swimming throughout the year which may negate this.

The model also does not account for cumulative swimming events over multiple days but rather risk of illness from swimming activities on a given day. Colder parts of the country may

have less days with warmer water, but when it is warmer, the exposure times may be no different. Anecdotally people may also adapt to local conditions, with people in colder regions swimming in water that is colder, than those in warmer regions. As noted above there is no New Zealand specific data to support running different models for different locations. Local data would be required to be collected to support the use of different models to those proposed by this study. If this type of data was obtained, the model could be rerun for different age groups if required to address risk characterisation questions or for specific sites.

16.4.3 Pathogen and indicator concentrations

Internationally pathogen concentrations for a freshwater QMRA are either (i) based on data collected directly from sites (Skiendzielewski et al, 2024) or (ii) are based on the concentrations of pathogens and indicators in raw faecal matter or from sampling where pathogens enter the water and are adjusted to account for dilution or relative survival/die off rates of pathogens and indicators (Ahmed et al, 2024; Soller et al, 2010 and 2014; Boehm and Soller, 2020).

This study collected both indicator and pathogen data directly at recreational water sites from the same water sampling session. Providing good quality data for investigation of relationships between the pathogens and indicators including the variability inherent in such relationships (Figures 17, 20, and 22). This approach removes the need for assumptions about dilution or the relative die off rates for pathogens and indicators.

To support national level metrics, the indicator and pathogen relationships used in the QMRA model have been derived from data taken from sites across the country, for sites with different combinations of faecal sources and under a range of weather conditions and seasons.

16.4.4 Dose-response

The translation of exposure dose of a pathogen to infection or illness is governed by if a strain of a micro-organism will cause illness (the dose-response relationships as discussed in sections 13.2.3 and appendix E). The dose response has components of:

- Variability in how a given dose of pathogen infects or causes symptoms between individuals and also for different population groups.
- Uncertainty in dose-response relationships and model parameters given the data that is available to develop the relationships. This includes a data gap relating to low doses most likely to be encountered by swimmers.

Dose-response is often the largest source of uncertainty in QMRA outputs, as acknowledged internationally (WHO, 2016, USEPA, 2024). For example, applying the Teunis et al. (2018) *Campylobacter* dose response model that includes uncertainty in the model parameters in the QMRA, gives a median estimate of 6% swims resulting in illness, with a 95% uncertainty interval or 0.5 to 30% (Appendix E, Table 60).

For ethical reasons, dose response relationships are not likely to improve in the foreseeable future, so there will always be uncertainty about the exact health outcomes. Which leaves two approaches to using the QMRA outputs:

(i) Report and compare the dose as well as health outcomes.

(ii) Keep the dose-response consistent across QMRA to allow effects of other important variables or interventions to be compared and fair comparison with guideline values.

In line with the above recommendation to keep dose-response relationships consistent for freshwater QMRA, the campylobacteriosis dose-response used in the 2002 QMRA model has been used for the 2025 simulation model. If further dose-response or epidemiology data becomes available in the future, the dose-response aspect of the simulation models should be re-evaluated.

There was interest by the project's governance and stakeholder groups to understand the different rates of infection and illness associated with different population groups, such as the very young, pension aged, immunocompromised or different ethnic groups. There are no published dose-response relationships or specific data for these population groupings for the bacterial or protozoa pathogens considered in this study. Therefore, there is no evidence to support QMRA analyses for population subgroups at the current time and for ethical reasons this is unlikely to change in the foreseeable future.

16.5 E. COLI AND ENTEROCOCCI AS HUMAN HEALTH RISK INDICATORS

The statistical analyses in section 8 indicated that *E. coli* and enterococci were positively associated with the prevalence of *Campylobacter Salmonella* and STEC, and weakly so with *Cryptosporidium* and *Giardia*. The *E. coli* and enterococci concentrations were also positively associated with *Campylobacter* concentrations (section 9). Both of these faecal indicator bacteria were detected in most samples tested, and across a useful range of concentrations. Quantile regression analysis was used to generate distributions of *Campylobacter*, *Cryptosporidium* and *Giardia* concentrations for a given *E. coli* or enterococci concentration to be used in the QMRA.

The QMRA simulation model outputs revealed that either *E. coli* or enterococci can be used in the New Zealand context to define changes in the percentage of freshwater swims resulting in campylobacteriosis but are not suitable for use for cryptosporidiosis or giardiasis.

The estimated percentage of swims in water with different *E. coli* or enterococci concentrations that would result in campylobacteriosis infection, illness and Disability adjusted life years (DALY) per 1,000 swims are presented in Table 54.

Table 54: Estimated percentage of swims in water with different *E. coli* and enterococci concentrations resulting in campylobacteriosis infection, illness and DALYs per 1,000 swims. Based on all data collected.

MPN / 100 ml	Log_{10} /100 ml	Percentage of swims resulting in infection with *Campylobacter*.		Percentage of swims resulting in illness (campylobacteriosis)		DALY due to campylobacteriosis per 1,000 swims	
		E. coli	Enterococci	*E. coli*	Enterococci	*E. coli*	Enterococci
40	1.6	-	4.5	-	2.0	-	0.3
50	1.7	2.0	-	1.0	-	0.1	-
100	2.0	3.0	6.0	1.0	2.5	0.2	0.4
200	2.3	4.5	7.5	2.0	3.0	0.3	0.5
300	2.5	6.0	8.5	2.5	3.5	0.4	0.5
400	2.6	7.0	9.0	3.0	3.5	0.4	0.5
500	2.7	8.0	9.5	3.0	4.0	0.5	0.6
600	2.8	9.0	10.5	3.5	4.0	0.5	0.6
800	2.9	9.5	12.0	4.0	5.0	0.6	0.7
1000	3.0	9.5	12.5	4.0	5.0	0.6	0.8
130	2.1	[0.0][a] 3.5		1.5		0.2	
260	2.4	[0.9] 5.5		2.0		0.3	
540	2.7	[4.9] 8.5		3.5		0.5	

Notes: a: [ ] Percentage of swims resulting in *Campylobacter* infection as estimated by the 2002 model (McBride et al., 2002). In consideration of the modelling uncertainties, the percentage of swims estimated to result in infection are rounded to nearest multiple of 0.5%, and DALY estimates are rounded to one decimal place.

Even at low *E. coli* (50 MPN/100 ml) or enterococci (40 MPN/100 ml) concentrations, there remains a low risk of campylobacteriosis (1-2% of swims resulting in illness). This contrasts with the 2022 FMRP study where a zero risk was determined for low *E. coli* concentrations. The difference in results is due to the difference in modelling approach and the reality that zero risk is not achievable in open air freshwater (Kay et al, 2006). There will always be a risk of illness from swimming in water with the potential for faecal inputs.

For *E. coli* concentrations above 1,000 MPN/100 ml there was not a significant trend of changing *Campylobacter* concentrations with increasing *E. coli* concentration (section 11.3.2). This doesn't mean that higher levels of *E. coli* won't have higher risks of illness, just that based on the data collected during this project, the *Campylobacter* risk levels off.

The levelling off for gastroenteritis illness incidence as water *E. coli* and enterococci concentration increases was also observed in the freshwater study by Wiedenmann et al. (2006). While this study was not pathogen specific and was for a swimming activity of 10 minutes, it still suggests the reduction in rate of increase in risk occurs as the indicators increase (Table 55Table 55). The data in the table is grouped by the quartiles of the microorganism concentration measured in the study.

Table 55: Gastroenteritis rates observed in freshwater swimming trials in water with different *E. coli* or enterococci concentrations adapted from Wiedenmann et al (2006).

Parameter	Microorganisms /100 ml	Illness cases	Participants	Incidence rate (%)
Unexposed at beach	-	26	921	2.8
E. coli	≤ 72	4	207	1.9
	73 to181	11	212	5.2
	182 to 379	14	211	6.6
	380 to 4,600	17	208	8.2
Intestinal enterococci	≤ 14	5	208	2.4
	53	9	212	4.2
	101	14	210	6.7
	1,190	18	208	8.7

However, higher concentrations of *E. coli* are likely to indicate higher health risks for other pathogens. For MST markers, 96% of samples with greater than 10,000 *E. coli*/100 ml, had high concentrations of human or ruminant markers. This drops to 90% for samples with 5-10,000 *E. coli*, 85% for samples with 1-5,000 *E. coli* and 47% for those with less than 1,000 *E. coli*. For the human viral indicator CrAssphage, 92% of the samples with CrAssphage levels with infection risks >10% were in samples with greater than 1,000 *E. coli*/100 ml, and 35% of these had levels greater than 10,000 *E. coli*/100 ml. While below 1,000 *E. coli*/100 ml there are clear risk reductions, above this level the risk is clearly high. We would suggest that all rivers with ongoing samples greater than 1,000 *E. coli*/100 ml should require further investigation with prioritisation of investigations on the basis of increasing *E. coli* concentrations above this 1,000 threshold. Overseas guidelines do not include thresholds above 1,000 *E. coli*/100 ml (McBride and Soller, 2017, Leonard and Eaton, 2021, Leonard, 2025).

The WHO Guidelines on recreational water quality (WHO, 2021) recommends using enterococci to set national health-based targets for recreational marine and freshwater bodies. The rationale for that recommendation is that (i) “gastrointestinal illness occurs at a higher rate in seawater than in fresh water at a given FIB level”, (ii) “No statistical relationship has been established for *E. coli* that can support a dose-response guideline value” and (iii) the use of two FIB introduces avoidable complexity.

The health-based targets were based on trials in UK marine environment and followed swimmers who took part in a 10 minute swimming activity including three head immersions (Kay et al., 1994). These trials measured water enterococci concentrations (maximum of 158/100 ml), but did not measure *E. coli* in the water. The later German freshwater trial (Wiedenmann et al, 2006) found increasing risk of illness with increasing *E. coli* or enterococci (Table 55). The relative illness rates of these studies have been used to justify that illness rates for given enterococci concentrations are greater in marine compared with freshwater, however this may have been due to differences in the susceptibility of the different populations (WHO, 2009) or different pathogens causing the illnesses.

Comparisons of New Zealand Recreational Water Guidelines with overseas approaches have been undertaken previously (McBride and Soller, 2017, Leonard and Eaton, 2021, Leonard, 2025). These reviews included guidelines from WHO (2021), Australia (2008), EU (2006), US EPA (2012) and Canada (2012). In this study, illness as the endpoint for health outcomes aligns more closely with international approaches. While there is variation in the statistical metrics used in guidelines, the findings in this study do align with international values.

16.6 SITE SPECIFIC QMRA

Components of the current national recreation water QMRA can be re-used to design site specific QMRA models. These include the water ingestion volumes, dose-response models and possible pathogen concentrations given FIB concentrations, where pathogen sampling at a site has not been undertaken.

Regression modelling of the study data has shown between site variation in both the prevalence of detectable *Campylobacter* (section 8.8) and the *Campylobacter* concentration (section 9.4) above that described by the variables identified by the model using the complete dataset. MST marker profiles and environmental variables such as rainfall prior to water use, turbidity, water temperature, dissolved oxygen and pH may all have site specific impacts on *Campylobacter* transport and survival and should be considered for site specific risk assessment.

The Dutch implementation of QMRA for demonstrating microbiologically safe drinking water (Box 9.1 of WHO, 2016) identified a number of lessons which should be considered. To create a level playing field for different councils, it is recommended that any guidelines for use of recreational water site specific QMRA in New Zealand should require specified volumes of water ingested, dose-response relationships and minimum requirements for pathogen or indicator sampling.

16.7 CONCLUSIONS

This study provides an improved and expanded data set describing the relationships in New Zealand rivers between a range of microbial indicators and pathogens and an updated simulation model for estimating the risk to swimmers for gastrointestinal pathogens. Consistent with the previous FMRP study, the risk of infection following exposure to river water can be estimated based on concentrations of *E. coli* and the risk of *Campylobacter* infection. The percentage of swims resulting in infection, is predicted to be slightly higher than predicted in the 2002 model, for a given *E. coli* concentration. This current modelling extends the analysis to risk of illness which is more in line with international approaches.

Both *E. coli* and enterococci are useful indicators of faecal contamination and of potential health risk. Increasing concentrations of these faecal indicator bacteria are associated with increasing risk of faecally-associated pathogens.

A risk of illness is present across all rivers, and all sources of contamination. Wild birds contribute faecal material to most rivers and create a background risk of illness beyond any anthropogenic inputs. Anthropogenic sources do, however, increase that risk. Ruminant faecal sources increase the likelihood of *Campylobacter* and STEC, while human faecal sources increase the likelihood of *Campylobacter* and *Salmonella*. Human faecal matter is

also the source of human infective viruses which may need to be considered when evaluating a site.

Elevated levels of faecal indicators should prompt the use of MST tools to understand potential sources of contamination. In the case of human sources being identified, the concentrations of the MST marker, CrAssphage, can be used to estimate potential norovirus risk. For non-human sources, existing New Zealand datasets are insufficient to support differential illness risk based on faecal source. MST also provides guidance for water managers for mitigation opportunities which can support water quality improvements.

Recreational water guidelines are a science-informed policy decision. Analytical methodologies, pathogen carriage, land management practices, and QMRA modelling approaches have all changed since the FMRP study carried out over 20 years ago, and those changes are reflected in the conclusions of this current study. Revision of New Zealand's recreational water quality guidelines also needs to include revisiting sampling, statistical analysis, responses to elevated detections, and alignment with international guidelines. The data and QMRA modelling generated in this study allows for implications of any adjustments to recreational guidelines to be assessed against health risks, using science that is both current, and New Zealand specific.

APPENDIX A: HEALTH EFFECTS

A.1 **Campylobacteriosis**

Campylobacteriosis is the most frequently notified disease in New Zealand, with 6,089 cases notified in 2023 (116.6 cases per 100,000 population) and 989 hospitalisations (Armstrong et al, 2024). There is marked seasonality in notifications, with the peak in spring and summer. The rate of campylobacteriosis is generally higher among males than females, with rates of 129.7 and 103.5 per 100,000 respectively in 2023 (Armstrong et al, 2024). The highest age-specific notification rate is in the 1 to 4 years age group (206.4 per 100,000 in 2023).

Traditionally, campylobacteriosis has mainly been attributed to *Campylobacter jejuni* and *C. coli*, but other species such as *C. fetus* are increasingly recognised as human pathogens.

Campylobacteriosis is characterised by muscle pain, headache and fever followed by self-limiting watery or bloody diarrhoea, abdominal pain and nausea.[9] Symptoms typically last for 3-7 days.

A.1.1 Sequelae

In a small proportion of cases, the characteristic acute gastrointestinal illness (AGI) due to *Campylobacter* spp. can be followed by more long-lasting illnesses. The sequelae most commonly associated with campylobacteriosis are Guillain-Barré syndrome (GBS), reactive arthritis (ReA) and irritable bowel syndrome (IBS).

Guillain-Barré syndrome

Guillain-Barré syndrome is the most common cause of acute flaccid paralysis in polio-free regions (Tam et al, 2007). It is characterised by an autoimmune pathology, with antibodies produced against pathogenic organisms, such as *Campylobacter* spp., cross-reacting with nerve ending antigens, resulting in neurological damage (Tam et al, 2007).

The first symptoms of GBS are pain, numbness, paraesthesia (‘pins and needles’) and weakness in the limbs (Hughes and Cornblath, 2005). The disease usually reaches its most serious stage after 2-4 weeks. After a plateau phase, recovery begins but the rate and extent of recovery are variable, with approximately 20% of cases still disabled after a year (Hughes and Cornblath, 2005). In severe cases patients may be unable to walk, and in about 25% of cases, weakness of the respiratory muscles necessitates artificial ventilation. In general, between 4 – 15% of cases are fatal (Hughes and Cornblath, 2005).

Studies have suggested that GBS occurs in 2-3 of every 10,000 cases of campylobacteriosis (McCarthy and Giesecke, 2001; Tam et al, 2006). The follow-up investigation to the large waterborne campylobacteriosis outbreak in Havelock North identified three subsequent GBS cases and estimated a total of 7570 campylobacteriosis cases – a rate of 4.0 per 10,000 campylobacteriosis cases (Gilpin et al, 2020). However, due to the small number of GBS

[9] https://www.mpi.govt.nz/dmsdocument/21548/direct Accessed 19 September 2024

cases the 95th percentile confidence interval associated with this estimate is quite large (1 - 12 per 10,000).

Campylobacteriosis-related GBS cases can also be estimated by examining GBS cases for evidence of a preceding *Campylobacter* infection. This approach has been more commonly employed internationally, with estimates ranging from 13 – 61% of GBS cases associated with a preceding *Campylobacter* infection (Boucquey et al, 1991; Caudie et al, 2011; Jackson et al, 2014; Jacobs et al, 1998; Koga et al, 2001; Mishu and Blaser, 1993; Mishu et al, 1993; Rees et al, 1995; Sharma et al, 2011; Tam et al, 2007). Baker et al (2012) analysed changes in GBS hospitalisation in New Zealand following a 52% decrease in campylobacteriosis notifications and concluded that GBS hospitalisations had decreased by 13%. This suggests that 25% of GBS hospitalisations in New Zealand are campylobacteriosis-related, equating to 35 campylobacteriosis-related GBS cases in 2023.

Reactive arthritis

Reactive arthritis (ReA) refers to inflammatory arthritis occurring subsequent to a gastrointestinal or genitourinary infection (Pogreba-Brown et al, 2021). Joint pain usually starts within two weeks of the inciting infection and typically lasts for 3-12 months, with a small proportion of cases developing chronic disease. A review of ReA cases in southern New Zealand found that 37% were hospitalised at some stage and almost half experienced a disease course longer than two years (Highton and Priest, 1996).

Reactive arthritis has been estimated to affect 1 – 26% of campylobacteriosis cases (Ajene et al, 2013; Hannu et al, 2002; Keithlin et al, 2014; Schönberg-Norio et al, 2010). Walker et al (2022) provided New Zealand specific information with their follow-up study to the 2016 Havelock North campylobacteriosis outbreak. This study involved an initial telephone survey, followed by telephone interviews with a rheumatologist. On the basis of their campylobacteriosis status, respondents were classified as culture confirmed (CC, 106 respondents), probable campylobacteriosis (PC, 47 respondents) or no diarrhoea (ND, 113 respondents). Following the rheumatologist interview, 19 CC, 4 PC and 2 ND cases were classified as probable ReA cases. Due to non-responses at each stage, the study calculated the incidence of ReA in different manners. Expressing the number of probable ReA cases as a proportion of the total original cohort (minimum incidence) gives incidence rates of 8.5, 2.8 and 0.6% for the CC, PC and ND groups, respectively. Assuming those who reported symptoms in the screening interview but did not proceed to the rheumatologist interview would have been as likely to be diagnosed with ReA as those who did complete the rheumatologist interview and using the number of respondents who completed the screening interview as the denominator, incidence rates of 23.9, 12.4 and 2.2% can be calculated. It is uncertain to what extent those who had suffered post-campylobacteriosis arthritic symptoms would have been motivated to participate in this study compared to cases who had not experienced symptoms.

Irritable Bowel Syndrome

Following acute bacterial gastrointestinal illness, up to one-third of cases may experience prolonged gastrointestinal symptoms, which in some cases will meet the clinical criteria for post-infectious IBS (PI-IBS) (DuPont, 2008). Irritable bowel syndrome is a chronic episodic condition characterised by abdominal discomfort, pain and altered bowel habits (DuPont, 2008; Scallan Walter et al, 2019).

While related to shigellosis, rather than campylobacteriosis, a cohort of 124 Korean hospital workers infected with *Shigella* due to contaminated food in the hospital employee cafeteria were interviewed at 1, 3, 5, 8, and 10 years after the outbreak, which occurred in 2001 (Youn et al, 2016). Twelve cases who reported IBS one year after the outbreak were assessed at each subsequent time point. After three years, nine of the 12 cases still reported symptoms of PI-IBS. This decreased to five cases at five years and two cases at eight years, before increasing to four cases at 10 years. Drawing on data from the same Korean outbreak, Haagsma et al (2010) assigned a mean duration of 5 years in assessing the burden of PI-IBS.

An European study collected information from IBS cases attending medical clinics by means of an internet-administered questionnaire (Card et al, 2018). Of 7552 respondents, 1004 met the study definition of PI-IBS. After one year, 19.7% of PI-IBS cases had recovered. Unfortunately, no longer follow-up was conducted.

Three systematic reviews and meta-analyses that considered the occurrence of PI-IBS subsequent to primary infection with specific pathogens were identified. A meta-analysis by Keithlin et al (2014) estimated that approximately 4% of campylobacteriosis cases will subsequently develop IBS. In contrast, a meta-analysis by Porcari et al (2024) concluded that 20.7% of those with *Campylobacter* infections will develop PI-IBS. Svendsen et al (2019) calculated a pooled prevalence of PI-IBS after *Campylobacter* infection of 12%.

Scallan Walter et al (2019) used medical records to identify a cohort of campylobacteriosis cases (*n* = 4143). Of these cases, 163 (3.9%) were diagnosed with IBS over the study period of four years, with 2.7% diagnosed within the first year after infection. Given the long duration of this study there are some doubts as to whether all the cases of IBS were PI-IBS and as such, 3.9% should be viewed as an upper limit for incidence of PI-IBS.

In a 2006 New Zealand study (Moss-Morris and Spence, 2006), 592 patients who tested positive for *Campylobacter*, but without existing bowel disease, were followed up three and six months after their initial infection. At three months, 83 campylobacteriosis cases (15%) reported symptoms consistent with IBS. By six months, this figure had decreased to 59 (11%). It is important to note that cases were not examined by a clinician.

A study conducted in Minnesota carried out six- and nine-month follow-ups to laboratory-confirmed cases of campylobacteriosis (Berumen et al, 2021). Of 1667 respondents, 249 (14.9%) had IBS prior to their *Campylobacter* infection. Of the remaining 1418 cases, 301 (21%) developed symptoms that met the clinical criteria for IBS subsequent to their *Campylobacter* infection. Of the PI-IBS cases, 28% were assessed as mild, 47% as moderate and 18% as severe.

A.1.2 Burden of disease

The most recent estimate of the burden of campylobacteriosis in New Zealand was derived for the 2013 year (Cressey et al, 2014). It was estimated that 65,200 cases of campylobacteriosis occurred during the year, resulting in 32 cases of GBS and 676 cases of ReA. Irritable bowel syndrome was not included as an outcome of campylobacteriosis at that time. The overall burden of campylobacteriosis was estimated to be 976 DALYs, the highest of any of the bacterial enteric diseases considered. The overall DALYs per case for

campylobacteriosis in 2013 was 0.015. This metric provides an index of the frequency and severity of adverse health effects associated with the organism. For comparison, although the total DALY burden due to norovirus infection was estimated to be substantially higher than for campylobacteriosis, the DALYs per case for norovirus infection was estimated to be lower, at 0.010.

A.2 **Salmonellosis**

Salmonellosis generally refers to infections due to genotypes of *Salmonella enterica* subsp. *enterica* but does not include infections caused by *S. typhi* and *S. paratyphi*. Most cases of salmonellosis are mild but infections can be life threatening[10]. Symptoms include rapid onset of fever, nausea, abdominal pain, diarrhoea and in some cases vomiting.

New Zealand's rate of salmonellosis is modest by international standards, with 827 notified cases (15.8 cases per 100,000 population) in 2023 (Armstrong et al, 2024). Salmonellosis results in a reasonably high rate of hospitalisation, with 218 cases (26%) hospitalised in 2023. There is evidence that a substantial proportion of cases are acquired overseas, with 256 notified cases in 2023 reporting that they had travelled overseas during the incubation period of the organism. This is supported by the observation that salmonellosis cases decreased from approximately 1200 in 2019 to less than 800 in 2020, likely linked to the lack of travel during the COVID-19 pandemic. The rate of salmonellosis has still not returned to pre-COVID-19 levels.

The rate of salmonellosis is generally similar between males and females (16.3 and 15.3 cases per 100,000 population, respectively in 2023) (Armstrong et al, 2024). The highest age-specific rates are amongst those aged < 5 years. While rates of salmonellosis in New Zealand in 2023 were highest amongst those of European or other ethnicity, rates of hospitalisation due to salmonellosis were highest amongst those of Māori or Pacific peoples ethnicity (Armstrong et al, 2024).

A.2.3 Sequelae

While the 2015 WHO global burden of foodborne disease study did not include any sequelae in the estimation of disease burden due to salmonellosis, other studies have recognised ReA and occasionally IBS and/or inflammatory bowel disease (IBD) as sequelae to salmonellosis (Chen et al, 2012; Gibney et al, 2014; Haagsma et al, 2010; Kemmeren et al, 2006; Majowicz et al, 2020; Mangen et al, 2015; Pires et al, 2019).

Reactive arthritis

A systematic review and meta-analysis of 17 studies by Pogreba-Brown et al (2021) found an association between ReA and salmonellosis They estimated that 3.9% of reported salmonellosis cases developed ReA, with estimates from individual studies ranging from 0.1 to 62.5%. However, for studies in which ReA was diagnosed by a specialist, the range was

[10] https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) Accessed 15 April 2025

substantially narrower (1.3 to 9.0%). A study of 60 ReA cases in southern New Zealand found evidence of prior *Salmonella* infection in two cases (Highton and Priest, 1996).

Irritable bowel syndrome

A study conducted in Germany concluded that PI-IBS was more likely to occur after infection with *Campylobacter* than *Salmonella* (Schwille-Kiuntke et al, 2011). However, PI-IBS symptoms following a *Salmonella* infection appeared to be more severe. In a more recent meta-analysis, a pooled prevalence of 12% was estimated for IBS subsequent to salmonellosis (Svendsen et al, 2019). This is similar to the prevalence (11.7%) reported by Mearin et al (2005) in their follow-up to a large salmonellosis outbreak.

A.2.4 Burden of disease

The most recent estimate of the burden of salmonellosis in New Zealand was derived for 2013 (Cressey et al, 2014). It was estimated that 6,820 cases of salmonellosis occurred during the year, resulting in 145 cases of ReA and 4 cases of IBD. The overall burden of salmonellosis was estimated to be 121 DALYs, with most of the burden due to the primary gastroenteritis. The overall DALYs per case for salmonellosis in 2013 was 0.018, similar to that for campylobacteriosis.

A.3 **Shiga toxin-producing *E. coli* infection**

New Zealand has one of the highest reported rates of STEC infection in the world. In 2023, 1006 STEC infection cases were notified in New Zealand (19.3 cases per 100,000 population), with 235 cases (23.4%) reported to have been hospitalised (Armstrong et al, 2024). Annual STEC infection notifications increased from < 200 cases in 2014 to approximately 1,000 cases by 2019. However, it appears that much of this increase is due to changes in testing methods. There is little difference in rates of STEC infection between males and females (18.9 and 19.6 cases per 100,000 population respectively in 2023). The highest age-specific rates of STEC infection are in those aged less than five years of age.

Symptoms of STEC infection range from mild watery diarrhoea, abdominal cramps, fever and vomiting to bloody diarrhoea, thrombocytopaenia, uraemia and acute kidney failure[11]. Both O157 and non-O157 STEC have similar clinical presentations, but disease severity can differ depending on the serotype.[12] Primary STEC infections are often characterised by haemorrhagic colitis (bloody diarrhoea). In 2023, this information was reported for 693 STEC infection cases, with 208 (30%) reported with haemorrhagic colitis and 70% without (Peter Cressey, ESR, personal communication).

A.3.1 Sequelae

The most well-established sequel to STEC infection is haemolytic uraemic syndrome (HUS), particularly amongst the young (Petruzziello-Pellegrini and Marsden, 2012). This syndrome is characterised by thrombocytopenia (low platelet levels in the blood), nonimmune

[11] https://www.who.int/news-room/fact-sheets/detail/e-coli Accessed 15 April 2025
[12] https://www.mpi.govt.nz/dmsdocument/26030-Shiga-toxin-producing-Escherichia-coli-STEC Accessed 24 September 2024

haemolytic anaemia (premature death of red blood cells), and acute renal failure (Petruzziello-Pellegrini and Marsden, 2012).

In New Zealand, part of the case definition for STEC infection is “any case of haemolytic uraemic syndrome (HUS) or thrombotic thrombocytopenic purpura (TTP) with or without a history of an acute onset diarrhoeal illness”[13]. In other words, HUS is considered to always be a sequel to a STEC infection. For a proportion of HUS cases the condition will progress to end-stage renal disease (ESRD).

Haemolytic uraemic syndrome

In 2023, there were 36 incident hospitalised cases of HUS in New Zealand, with 26 patients (72%) being less than five years old (Armstrong et al, 2024). Of those patients under five years of age, 15% developed HUS, while 2.5% of cases 5-14 years and 1.1% of cases 15+ years developed HUS (Armstrong et al, 2024).

In 2023, Wong et al. published results of a 23-year study of diarrhoea-associated HUS in young children in New Zealand. This study included 226 cases with a median age at presentation of 2.8 years (Wong et al, 2023). The majority of cases (148; 74.7%) were referred to the paediatric nephrology centre in Auckland, with 127 (56.2%) undergoing dialysis. The median duration of dialysis was 9 days (range 1-38 days), with all dialysed cases recovering sufficient kidney function to be dialysis independent. The authors noted that 34% of cases required dialysis for 5-9 days, 28% from 10 to 14 days and 20% for 15 days of more. One patient developed kidney failure. Seizures occurred in 31 cases (13.7%), all within 48 hours of HUS diagnosis. Multiple and complex seizures developed in 11 children with 3 fatalities (1.3% of HUS cases).

End stage renal disease

A proportion of HUS cases will develop ESRD and require ongoing dialysis and/or kidney transplantation. A 20-year study in the US reported that 3% of post-diarrhoeal HUS cases develop ESRD (Siegler et al, 1994). Tang et al (2012) reviewed the course of HUS-associated ESRD in Australia and New Zealand. They reported that first treatments for HUS-associated ESRD cases included haemodialysis (68%), peritoneal dialysis (29%) and direct renal transplant (4%). For cases on dialysis, 29% died, with a median time to death of 6.3 years. Renal transplants were received by 54% of HUS-associated ESRD cases, with a median time to the first allograft of 1.4 years. Renal transplants often fail, with only 23% of allografts still surviving after 10 years. While ESRD cases may receive subsequent renal transplants following failure of the allograft, only 19% received a second transplant, 2% received a third and only one case out of the study cohort received a fourth allograft. Only 9% of HUS-associated ESRD cases recovered renal function and only 29% of these cases never returned to renal replacement therapy (dialysis or renal transplant).

A.3.2 Burden of disease

The most recent estimate of the burden of STEC infection in New Zealand was derived for the 2013 year (Cressey et al, 2014). It was estimated that 3,800 cases of STEC infection

[13] https://www.tewhatuora.govt.nz/for-health-professionals/clinical-guidance/communicable-disease-control-manual/verocytotoxin-or-shiga-toxin-producing-escherichia-coli-vtecstec Accessed 15 April 2025

occurred during 2013, resulting in 82 cases of HUS and 9 cases of ESRD. The overall burden of STEC infection was estimated to be 510 DALYs. The major contributors to this DALY estimate are the large number of years of life lost (YLL) due to mortality amongst young HUS cases and the burden of both mortality and morbidity associated with ESRD. The overall DALYs per case for STEC infection in 2013 was 0.13, approximately 9-fold higher than for campylobacteriosis.

A.4 **Cryptosporidiosis**

Cryptosporidiosis is a parasitic infection caused by protozoa from the genus *Cryptosporidium*. In New Zealand, the most common species associated with human infections are *C. hominis* and *C. parvum* (Garcia-R et al, 2020). Symptoms of cryptosporidiosis most commonly include stomach cramps and watery, foul-smelling diarrhoea.[14] Other symptoms may include fever, nausea, vomiting, loss of appetite and weight loss.

The rate of notified cryptosporidiosis cases in New Zealand is moderately high by international standards, with 831 cases notified in 2023 (15.9 cases per 100,000 population) (Armstrong et al, 2024). A moderate rate of hospitalisations was also noted, with 67 (8.1%) of cases in 2023 requiring hospitalisation (Armstrong et al, 2024). Where information on overseas travel was available, 11.5% of cases (60 out of 521) had travelled overseas during the incubation period of the organism.

Cryptosporidiosis notifications in New Zealand displayed no obvious trends between 2009 and 2023, with the exception of a spike in cases in 2018 (Armstrong et al, 2024). However, notifications did display a distinct seasonal pattern, peaking in spring. Notification rates in 2023 were slightly higher for females compared to males (16.6 and 15.2 cases per 100,000 population respectively), and age-specific rates were highest for children aged 1 – 4 years.

A.4.1 Sequelae

The WHO global burden of foodborne disease study did not include any sequelae in the estimation of the disease burden due to cryptosporidiosis (World Health Organization, 2015). However, a systematic review of long-term symptoms (up to 28 months) following *Cryptosporidium* infection found that persistent diarrhoea (25%), abdominal pain (25%), nausea (24%), fatigue (24%) and headache (21%) were common (Carter et al, 2020). While some studies have reported PI-IBS at rates up to 30% (Stiff et al, 2017), the evidence is less strong for PI-IBS as a sequel to cryptosporidiosis than for some other microbial diseases. Iglói et al (2018) reported the presence of joint pain in some cryptosporidiosis cases at follow-up, but symptoms were not clinically confirmed as ReA and the rate of symptoms was no greater than in the general population.

[14] https://info.health.nz/conditions-treatments/infectious-diseases/cryptosporidium Accessed 16 April 2025

A.4.2 Burden of disease

Cryptosporidiosis has not been included in previous New Zealand estimates of the burden of disease. However, an unpublished study by ESR estimated the total burden of cryptosporidiosis as 114 DALYs in the 2017 year (Cressey and Lake, unpublished). This estimate did not include any sequelae in the burden of cryptosporidiosis and the corresponding DALYs per case metric is similar to that for norovirus infection (0.012).

A.5 **Giardiasis**

Giardia spp., mainly *G. intestinalis*, are enteric protozoan pathogens. In 2023, 897 cases of giardiasis were notified in New Zealand (17.2 cases per 100,000 population), with 43 cases (4.8%) being hospitalised (Armstrong et al, 2024). Giardiasis case numbers were relatively steady at > 1,500 notified cases per year up to 2019 but have decreased substantially in subsequent years (Armstrong et al, 2024). There is little evidence of seasonality in giardiasis cases. Notification rates are generally higher in males compared to females (18.6 and 15.8 cases per 100,000 population in 2023, respectively) and are highest in those aged under five years, following by those aged between 30 – 39 years (Armstrong et al, 2024). It has been reported that only 49% of giardiasis cases are notified in New Zealand (Hoque et al, 2003). However, it is likely that the degree of under-reporting is even greater than this.

Giardiasis is characterised by abdominal cramps, nausea, acute or chronic foul-smelling diarrhoea, low-grade fever, anorexia and bloating.[15] Symptoms usually resolve within 7 to 10 days, but reinfection is common in endemic areas and chronic infections, lasting several months, can occur if untreated.

The WHO global burden of foodborne disease study estimated that 91% of giardiasis cases resulted in only mild diarrhoea (Kirk et al, 2015). This is consistent with a Canadian study that reported bloody diarrhoea in only 4 of 176 giardiasis cases (Ravel et al, 2013).

A.5.1 Sequelae

The WHO did not include any sequelae to giardiasis in its global burden of foodborne disease study (World Health Organization, 2015). Extensive follow-up of a large giardiasis outbreak in Norway identified higher prevalence of IBS and chronic fatigue amongst cases than unexposed controls (Litleskare et al, 2018). However, it has been suggested that the relationship between giardiasis and IBS may be more complex, with giardiasis exacerbating existing IBS (D'Anchino et al, 2002). *Giardia* has also been implicated as a triggering agent for ReA but this relationship appears less well established (Halliez and Buret, 2013).

A.5.2 Burden of disease

Giardiasis has not been included in previous New Zealand estimates of the burden of disease. An unpublished study by ESR estimated the total burden of giardiasis as 302 DALYs in the 2017 year (Cressey and Lake, unpublished). Given the substantial decrease in

[15] https://www.mpi.govt.nz/dmsdocument/11024-Giardia-intestinalis-Microbial-pathogen-data-sheet Accessed 23 September 2024

giardiasis notifications since that time, the disease burden is also likely to be substantially lower. The DALYs per case was estimated at 0.011 but this did not include any sequelae.

A.6 **Norovirus infection**

Norovirus infection is not a notifiable disease in New Zealand, and as such there is no systematic information available on the number of annual cases. However, since July 2000, Public Health Services in New Zealand have been encouraged to record all cases of gastroenteritis caused by non-notifiable or unknown food-borne intoxicants including those self-reported by the public. A causal organism is reported for a proportion of these cases, with norovirus being the most commonly reported organism.

In an Australian study, Gibney et al (2014) reported that norovirus infections accounted for 13.1% of all cases of AGI. A 2011 New Zealand study estimated that each New Zealander experiences 1.12 incidents of AGI per year (Adlam et al, 2011). Based on a New Zealand population estimate of 5,305,600 (as of 31 December 2023), these figures suggest that 778,438 cases of norovirus infection would have occurred in New Zealand in 2023, or approximately 14,700 cases per 100,000 population.

There are no well-characterised sequelae to norovirus infection.

A.6.1 Burden of disease

The most recent estimate of the burden of norovirus infection in New Zealand was derived for the 2013 year (Cressey et al, 2014). Based on rates from a British study (Tam et al, 2012), it was estimated that 211,000 cases of norovirus infection occurred during the year, resulting in 104 hospitalisations. The overall burden of norovirus infection was estimated to be 2195 DALYs, with the major contributor to this estimate being the large number of cases. The overall DALYs per case for norovirus infection in 2013 was 0.010.

APPENDIX B: METHODS

B.1 Culture methods and identification of bacteria by PCR

B.1.1 Total coliforms, *E. coli* and enterococci

Total coliforms *E. coli* and enterococci were analysed using Colilert Quanti-Tray and Enterolert assays (IDEXX Laboratories, Inc., Maine, US) (APHA, 2017a, 2017b) respectively. Samples were diluted 1:10 before analysis which gave a detection range of <10 MPN/100 ml to >24,200 MPN/100 ml. MPNs were calculated from the IDEXX Quanti-Tray 2000 MPN table.

B.1.2 Campylobacter

In phase 1, a 12 tube MPN analysis was undertaken to isolate *Campylobacter* spp. Freshwater water volumes of 1 x 700 ml, 2 x 100 ml, 3 x 1 ml, 3 x 0.1 ml, 3 x 0.01 ml were analysed giving a range of <0.11 MPN/100 ml up to >11,100 MPN/100 ml. In phases 2 and 3 the dilutions were 1 x 1000 ml, 3 x 10 ml 3 x 1 ml and 3 x 0.1 ml, giving a range of <0.096 to >1,100 MPN/100 ml.

The methods are described in (APHA, 2017c; Health Canada, 2014; International Organization for Standardization, 2017; MIMM, 2008). Briefly 100-, 700- and 1000-ml samples were filtered on 0.22 µm filters. Filters were placed in 25 ml of Bolton broth, except for those from 1000 ml sample which were placed in 100 ml of Bolton broth. Analysis of volumes 1 ml or less were directly added to 25 ml of broth. After incubation (41.5 +/- 1˚C in 10% CO_2, for 24 h), a loopful of broth was streaked onto modified Charcoal-Cefoperazone-Deoxycholate Agar (CCDA) plates. Putative positive colonies from the CCDA plates were restreaked onto Columbia Blood Agar (CBA) plates.

When colonies consistent with Campylobacter were observed on CCDA plates, the 1 ml of the broth which was the source of those colonies, was centrifuged and resuspended in 500 µl of 2% Chelex 100 Sodium form (Sigma Cat#C7901) in sterile Milli-Q (Millipore, Merck) water. Purified colonies from the CBA plates (1-2) were added into 500 µL of 2% Chelex 100 Sodium form (Sigma Cat#C7901) in sterile Milli-Q (Millipore, Merck) water.

The Chelex solution was heated at 95°C for 10 min, centrifuged 16,000 g for 5 min and the supernatant tested by conventional PCR using the *Campylobacter* multiplex assay of Wong et al (2004). Visualisation of the PCR amplicons was carried out on the MultiNA Microchip Electrophoresis System (Shimadzu) using fluorescence detection of the DNA products. Expected band sizes for the targets were 695 bp for *C. coli* (695 bp), 246 bp for thermotolerant *Campylobacter* and 99 bp for *C. jejuni*.

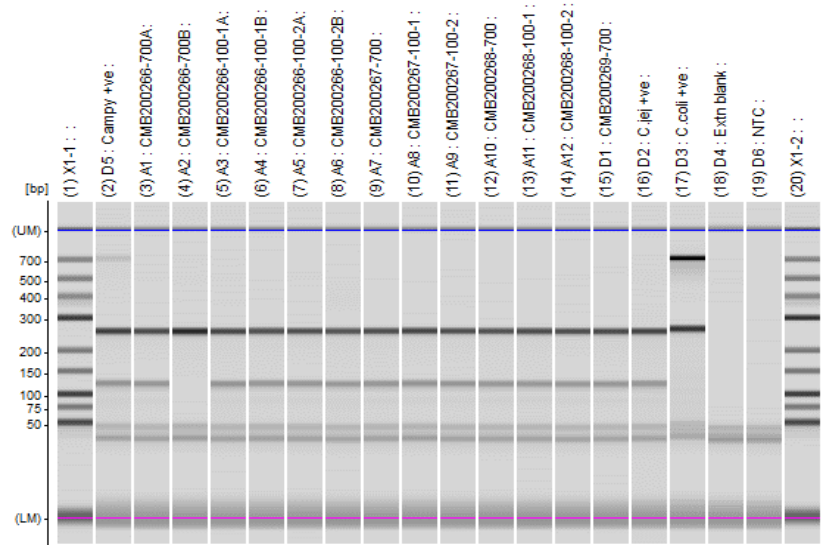


Figure 38: An example of a *Campylobacter* PCR detection using the MultiNA electrophoretic system. Note: From the top, the gene targets for *C. coli* (695 bp), Thermotolerant *Campylobacter* (246 bp) and *C. jejuni* (99 bp). There is a slight bp offset on the image

MPN values were calculated based on the results from testing of the broths, using Most Probable Number Calculator. Broths were noted as positive for *C. jejuni*, when both 246 bp thermotolerant *Campylobacter* band and 99 bp *C. jejuni* were detected; positive for *C. coli* when both 246 bp thermotolerant *Campylobacter* band and 695 bp *C. coli* bands were detected; and positive for thermotolerant *Campylobacter* when the 246 bp thermotolerant *Campylobacter* band was detected, with or without any specific *C. jejuni* or *C. coli* bands.

B.1.3 *Salmonella*

In phase 1 a 9 tube MPN analysis was undertaken to isolate *Salmonella* spp. Freshwater volumes of 1 x 700 ml, 2 x 100 ml, 3 x 1 ml, 3 x 0.1 ml were analysed. (APHA, 2017d; Public Health England, 2015). This provided a detection range of <0.11 MPN/100 ml up to >1,110 MPN/100 ml. In phases 2 and 3 presence/absence in 1 litre was determined from filtering 1 litre of sample.

Volumes greater than 1 ml were filtered through 0.45 µm filters and placed into 25 ml of buffered peptone water (BPW) broth. Volumes 1 ml or less were directly added to the broth. Broths were incubated at 37˚C for 18 h and then 1 ml of BPW was transferred into 10 ml of Muller-Kauffmann Tetrathionate Novobiocin Supplement broth, and 0.1 ml of BPW transferred into 10 ml of Rappaport-Vassiliadis Soya peptone broth, which were then incubated at 37˚C and 41.5˚C respectively for 24 h. A loopful of each broth was plated onto Xylose Lysine Deoxycholate and Hektoen Enteric agar and incubated for 24 h at 37˚C. Putative *Salmonella* colonies were plated on MacConkey agar, tryptic soy agar, inoculated into tryptone broth, urease broth and onto triple sugar iron agar or Lysine Iron Agar slopes. Isolates which were consistent with *Salmonella* then had polyO and polyH serology undertaken. If still indicative of *Salmonella*, biochemical testing was undertaken using Microgen or Microbact biochemical identification kits.

B.1.4 Shiga toxin-producing *E. coli*

A 9 tube MPN analysis was undertaken to isolate STEC (International Organization for Standardization, 2012). Freshwater volumes of 1 x 700 ml, 2 x 100 ml, 3 x 1 ml, 3 x 0.1 ml were analysed. In phases 2 and 3 presence/absence in 1 litre was determined from filtering 1 litre of sample.

Volumes greater than 1 ml were filtered through 0.45 µm filters and placed into 25 ml of modified trypticase soy broth with novobiocin. Volumes 1 ml or less were directly added to the broth. Broths were incubated at 37˚C for 24 h and then the broths tested by STEC multiplex PCR which detected *eae*, *hly*A, *stx*1 and *stx*2 genes (Paton and Paton, 1998). Broths that tested positive were plated on MacConkey Agar with Sorbitol, Cefixime, and Tellurite (CT-SMAC) and MacConkey STEC chrome agar, and incubated for 24 h at 37˚C. Twenty colonies from each plate were streak isolated and purified on CBA plates, and then multiplex PCR performed on pools of 5 isolates. If a pool was positive, then individual isolates were tested by multiplex PCR. This provides a detection range of <0.11 MPN/100 ml up to >1,110 MPN/100 ml.

B.2 Microscopy methods

Cryptosporidium and *Giardia* were analysed by MicroAquaTech, Massey University, using a IANZ accredited method (United States Environmental Protection Agency, 2005). Samples were filtered through a IDEXX Filta-Max® filter in the field in Phase 1 and 3 and 10 litres sent to the laboratory in Phase 2. Protozoa on the filter were eluted, the (oo)cysts pelleted and isolated using anti-*Cryptosporidium* and anti-*Giardia* immunomagnetic beads. The (oo)cysts and cysts are stained on well slides with a fluorescent label and DAPI[16]. The stained sampled was examined using UV fluorescence and differential interference contrast microscopy to confirm size, shape and fluorescent characteristics of *Cryptosporidium* and *Giardia*. Recovery was determined on 10 samples from different sites and ranged from 31 – 54% for *Cryptosporidium* and 22 – 49% for *Giardia*. The raw concentration data is reported as (oo)cysts/10 litres.

The volume tested was dependent on the volume which could be filtered in the field before the filter clogged. In Phase 1, the objective was to filter 100 litres, while in Phase 3 the objective was to filter 30 litres. High concentrations of suspended solids in some samples meant that smaller volumes of water were filtered for some samples.

For phases 1 and 3 the limit of detection for *Cryptosporidium* ranged from 0.11 – 20.8 oocysts/10 litres and for *Giardia* 0.11 – 10 cysts/ 10 litres depending on the volume filtered. In phase 2, samples (10 litres) were sent to directly to the laboratory for filtering giving a detection limit of <1 (oo)cyst/10 litres.

[16] DAPI 4',6-diamidino-2-phenylindole

B.3 Whole genome sequencing

Isolates for WGS analysis were grown on CBA and incubated at 37ºC for 24 h ± 2 h prior to DNA extraction using the Qiagen DNeasy kit. DNA quality and concentration was performed using PicoGreen (Quant-iT; Thermo Fisher Scientific). Sequencing libraries containing 1 ng of DNA were prepared using Nextera XT chemistry (Illumina, San Diego, CA. USA) for 150-bp pair-end sequencing on an Illumina NextSeq sequencer, according to the manufacture's recommendations (Illumina).

Initial sequence quality and species identification were determined using the Nullarbor version 2 pipeline (Seemann et al, n. d.). The multi-locus sequence type (ST) and genome assemblies were performed used within BioNumerics version 8.1 using Skesa v.2.3 (Applied Maths NV, Belgium). Assembly-based calls were performed with BioNumerics using the whole-genome MLST (wgMLST) schemes within BioNumerics.

The identified sequence types of *Campylobacter* were compared to New Zealand and international isolates available on PubMLST database (Accessed 1 December 2024) (Jolley et al, 2018), to determine if the sequence type (ST) had previously been associated with human cases of campylobacteriosis, or if it had previously been isolated from ruminant, poultry or wild birds. The literature was also scanned to identify any publications with ST assignments.

B.4 Quantitative PCR

B.4.5 Microbial source tracking markers, bacteria and protozoa

A maximum of two litres of freshwater was filtered through 0.45 µm cellulose ester membrane filters (Millipore, France). Due to high concentrations of suspended solids, it was not always possible to filter two litres of sample. Therefore, the limit of detection varied for each MST marker depending on the volume sampled. With volumes between 400 and 2000 ml filtered, a range of detection limits were determined ranging from <15 to <58 genetic copies (GC)/100 ml for GFD, <18 to <73 GC/100 ml for BacR, <15 to <61 GC/100 mL for HF183, <49 to <196 GC/100 ml for crAssphage and < 20 to <85 GC/100 ml for BiADO.

DNA was extracted from the filters using the PowerSoil Pro kit (Qiagen, Venlo, The Netherlands) protocol on the QiaCube extraction robot. Quantitative PCR analysis was undertaken on a LightCycler 480 (Roche) with each amplification performed in duplicate, except GenBac3 was performed in quadruplicate in phase 3. Probe assays were performed using a total assay volume of 20 µl, with 2 µl of DNA extract, 0.2 mg/ml of BSA and Perfecta Tough mix for single probes, or Perfecta Multi-Mix for duplexes. The qPCR targets, primers, probes and assay conditions for bacteria, MST markers (HF183, crAssphage, GenBac3 and BacR) and protozoa are given in Table 56. Initial denaturation was at 95°C for 10 minutes for *E. coli*, enterococci, protozoa, HF183 and crAssphage and for 5 minutes for all other bacteria. Cooling was 10 s at 40°C. Amplification conditions for each assay are given in Table 56.

BiADO, GFD were analysed by SYBR Green assays with a total volume of 25 µL including 2 µl of DNA extract and Roche SYBR Green 1 (Cat. 04 707 516 001, Roche Diagnostics Ltd). In phase 3, PerfeCta qPCR ToughMix was changed to Roche LightCycler 480® Master Mix Probe. The primers, concentrations and method are given in Table 57. The preincubation

period was 5 minutes at 95°C and the individual amplification conditions ae given in Table 57. Conditions for Melt curves used a pre-incubation period of 5 sec at 95°C, melt start at 65°C for 60 sec with continuous increase in temperature at increments of 0.11°C/s with a cooling period at 40°C for 10 s.

Each run included negative and positive controls, and gBlock or Ultramer standard curves. Master standard curves were used to determine the concentration. New curves were generated for HF183 – crAssphage, *Salmonella* invA, STEC, stx-1, stx-2 and enterococci (phase 1 and 2 data), *E. coli* (phase 3 data) based on the method of (Sivaganesan et al., 2010). Individual curves were used for *Salmonella* ttr and STEC eae, which were only in phase 1, and *Cryptosporidium parvum*, *C. hominus* and *Giardia lamblia* which were in phase 1 and 2.

Inhibition was tested by dilution, and interference was identified by Sketa salmon sperm (phases 1 and 2) or using internal controls based on low GenBac3 and inconsistencies in results from *E. coli* by qPCR and culture. This resulted in removal from the dataset of 42 samples. These 42 samples came from 15 sites and covered all observed land uses, with 50% of the samples from three sites, and 10 of those samples (24%) from a single site, which suggested a local contributing factor. The cause of the interference was not able to be identified. Kinzelman et al (2011) report interference in qPCR has been observed in inland water, particularly sites dominated by non-point source pollution.

Table 56: Bacteria and protozoa qPCR target organism or gene, primers and probes, and assay concentrations and thermocycling conditions for bacteria, protozoa and MST markers.

Micro-organism and target organism or gene	Primers and probes	Concentration (nmol/µl)	Amplification	References
E. coli 23S rDNA	F: 5'-GGTAGAGCACTGTTTtGGCA R: 5'-TGTCTCCCGTGATAACtTTCTC P: [6-HEX]-5'-TCATCCCGACTTACCAACCCG-BHQ	1000 1000 80	15 sec 95°C 60 sec 56°C	(Aw et al, 2019; Chern et al, 2011)
Enterococci 23S rDNA	F: GAGAAATTCCAAACGAACTTG R: CAGTGCTCTACCTCCATCATT P: 56-FAM/TGGTTCTCTCCGAAATAGCTTTAGGGCTA/3BHQ-1/	1000 1000 80	15 sec 95°C 60 sec 56°C	(Haugland et al, 2012; United States Environmental Protection Agency, 2012)
Campylobacter jejuni mapA	F: CTGGTGGTTTTGAAGCAAAGATT R: CAATACCAGTGTCTAAAGTGCGTTTAT P: 56-FAM/TTGAATTCCAACATCGCTAATGTATAAAAGCCCTTT/3MGBEc/	500 500 50	10 sec 95°C 20 sec 60°C 10 sec 72°C	(Best et al, 2003)
C. coli ceuE	F: AAGCTCTTATTGTTCTAACCAATTCTAACA R: TCATCCACAGCATTGATTCCTAA P: Hex/TTGGACCTCAATCTCGCTTTGGAATCATT/BHQ1	500 500 50		
Salmonella invA	F: TCGTCATTCCATTACCTACC R: AAACGTTGAAAAACTGAGGA P: 56-FAM/TCTGGTTGA/ZEN/TTTCCTGATCGCA/3IABkFQ/	900 900 100	15 sec 95°C 60 sec 55°C	(Hoorfar et al, 2000)
Salmonella ttr	F: CTCACCAGGAGATTACAACATGG R: AGCTCAGACCAAAAGTGACCATC	500 500	15 sec 95°C 60 sec 65°C	(Malorny et al, 2004)

Micro-organism and target organism or gene	Primers and probes	Concentration (nmol/µl)	Amplification	References
	P: 56-FAM/AAAGTCGGTCTCGCCGTCGGTG/BHQ-1/	50		
STEC STX-1	F: ACATTGTCTGGTGACAGTAGC R: CGACATTAAATCCAGATAAGAAGTAGT P: 56-FAM/ATCAGTCGT/ZEN/ACGGGGATGCAGATAAAT/3IABkFQ/	500 500 200	10 sec 95°C 20 sec 60°C 10 sec 72°C	(Derzelle et al, 2011)
STEC STX -2	F: ATGACAACGGACAGCAGTTAT R: CTGAACTCCATTAACGCCAGATA P: 56-FAM/ATGCAAATCAGTCGTCACTCACTGG /3BHQ-1/	500 500 200		
EHEC eae	F: GTAAGTTACACTATAAAAGCACCGTCG R: TCTGTGTGGATGGTAATAAATTTTTG P R: 56-FAM/AAATGGACATAGCATCAGCATAATAGGCTTGCT/3BHQ-1/	300 300 100	10 sec 95°C 20 sec 55°C 10 sec 72°C	Ibekwe et al (2004)
Cryptosporidium parvum 18S rDNA	F: CATGGATAACCGTGGTAAT R: TACCCTACCGTCTAAAGCTG P R: 56-VIC/ATCACATTAAATGT/BHQ1	400 400 100	15 sec 95°C 60 sec 54°C	(Mary et al, 2013)
C. hominus 18S rDNA	F: CATGGATAACCGTGGTAAT R: TACCCTACCGTCTAAAGCTG P: 56-FAM/ATCACAATTAATGT/BHQ1	400 400 100		
Giardia lamblia beta-giardin	F: GGCCCTCAAGAGCCTGAAC R: GGGCGATCGTCTCCTTCTC P R: 56-FAM/CTCGAGACAGGCATC/3MGBEc/	300 300 50	15 sec 95°C 60 sec 60°C	Baque et al (2011)
General Bacteriodales 16S rRNA (GenBac3)	F: GGGGTTCTGAGAGGAAGGT R: CCGTCATCCTTCACGCTACT P: 6-FAM-CAATATTCCTCACTGCTGCCTCCCGTA-TAMRA	500 500 100	10 sec 95°C 20 sec 60°C 10 sec 72°C	Shanks et al (2010)

Micro-organism and target organism or gene	Primers and probes	Concentration (nmol/µl)	Amplification	References
Bacteroidales 16S rRNA (BacR)	F: GCGTATCCAACCTTCCCG R: CATCCCCATCCGTTACCG P: FAM-CTTCCGAAAGGGAGATT-NFQ-MGB	500 500 100		Reischer et al (2006)
Human HF183	F: ATC ATG AGT TCA CAT GTC CG R: CTT CCT CTC AGA ACC CCT ATC C P: FAM-CTA ATG GAA CGC ATC CC-BHQ-1	600 600 300	15 secs 95°C 60 secs 60°C	(Ahmed et al, 2019)
Human CrAssphage CPQ_056	F: CAG AAG TAC AAA CTC CTA AAA AAC GTA GAG R: GAT GAC CAA TAA ACA AGC CAT TAG C P: Cy5-AAT AAC GAT TTA CGT GAT GTA AC/BHQ-2	600 600 300		

Table 57: qPCR target organism or gene, assay concentrations and specific amplifying conditions and methods for SYBR Green assays of MST markers.

MST marker	Primers	Concentration (nmol/µl)	Amplification	References
Human *Bifidobacterium adolescentis* (BiADO)	Primer A: CTCCAGTTGGATGCATGTC Primer B: TCCAGTTGACCGCATGGT Primer C:CGAAGGCTTGCTCCCAGT	250 250 250	10 sec 95°C 10 sec 60°C 20 sec 72°C	(Matsuki et al, 2004)
Wild bird Unclassified *Helicobacter* spp. 16S rRNA (GFD)	Primer A 5' TCG GCT GAG CAC TCT AGG G from Green et al Primer B 5' GCG TCT CTT TGT ACA TCC CA from Green et al	250 250	10 sec 95°C 10 sec 57°C 20 sec 72°C	Green et al (2012)

B.4.6 eDNA Analysis

Samples extracted for DNA analysis were sent to Wilderlab (www.wilderlab.co.nz) for Environmental DNA (eDNA) analysis, and results for mammals and wild birds (mt12S-V5 and mt16S) analysis are reported here. Assignment to species was undertaken using Wilderlab custom software (Wilkinson et al, 2018).

Identified counts reported by Wilderlab were converted to counts/100 ml based on the different volumes filtered.

B.4.7 Viruses

Samples (10 L) were sent by courier for overnight delivery to the ESR. Each water sample was seeded with 1 ml murine norovirus (10^6 plaque forming units (PFU)/ml) as a process control to calculate virus recovery rate and to evaluate potential RT-PCR inhibition.

Each 10 L water samples were first concentrated to 700-900 ml by hollow-fibre ultrafiltration (HFUF) method (Hill et al, 2005) with modifications. In brief, 1% w/v sodium polyphosphate was added to each water sample and filtered through the HFUF membrane at the flow rate of 150-300/min until a final volume of approximately 500 ml was achieved. To further concentrate viruses, polyethylene glycol 6000 precipitation (with a preceding centrifugation step and beef extract elution step where necessary) was performed as described elsewhere (Hewitt et al, 2007)). Finally, 2-20 ml concentrates were obtained from each 10 litres water sample and stored at -80℃ until viral extraction. The volume of concentrates varied with turbidity of water sample. In the beginning of the study, this was 2 ml for small pellet and 5 ml for larger pellets. With the exception of two samples, from 23 Feb 2020, all samples were suspended in 10 ml. Viral nucleic acid was extracted from 200 µl concentrates using Presto DNA/RNA Extraction Kit (Geneaid Biotech Ltd, Taiwan) with a minor modification (ie, 20 µl poly A solution was added after PCR inhibition removal step).

All nucleic acid samples were tested for potential RT-PCR inhibition. Those RT-PCR inhibitory samples were further processed using Onestep™ PCR Inhibition Removal Kit (Zymo, USA) and re-tested for inhibition. Virus recovery from each water sample was determined by comparing the concentration of murine norovirus to control nucleic acid sample (nucleic acid extracted murine norovirus spiked into similar volume of distilled water). Controls were included in each RNA extraction and RT-qPCR assay including 1000, 100, and 10 RTPCR units of norovirus GI and GII.

Previously published qPCR/ two-step RT-qPCR assays were used to detect adenovirus (Hernroth et al, 2002), enterovirus (Donaldson et al, 2002), norovirus genogroup I (GI) (Wolf et al, 2010), norovirus GII (Kageyama et al, 2003) and murine norovirus (Hewitt et al, 2011).

B.5 Numerical analyses

The statistical package R Version 4.4.1 (R Core Team, 2024) was used for plotting, statistical analysis and the QMRA simulation modelling.

B.5.1 Ingestion rate distribution fitting

The ingestion data of Dufour et al. (2017) was modelling using a distribution estimated by the ‘fitdistrplus’ R package (Delignette-Muller and Dutang, 2015). Candidate distributions with the property of only positive values were compared via the log likelihood values, Q-Q

and P-P plots. The chosen distribution for ingestion rate was a Weibull distribution [shape 1.15 (standard error 0.09), scale 57.0 (standard error 5.5)].

B.5.2 Quantile Regression

Quantile regression was performed using the R package 'qgam' (Fasiolo et al, 2021), which calculates non-parametric smoothed estimates of the required percentiles of the response variable given independent variables. The regression was applied to the $\log_{10}$ transformed concentration data.

Smoothing was implemented using thin plate regression splines and smoothing components were applied to both the estimated percentiles and the variance in the model fit. R formula of the form:

Model <- qgam (form =list(y ~ s(x, k=K, bs="tp"), ~s(x)), qu = P, data = Dataframe)

The number of knots (K) for fitting each percentile (P) was chosen based on the goodness of fit and reducing unexpected wiggle or overlapping of percentile lines.

Quantile regression models were examined for goodness of fit using the 'check' function of the 'qgam' package, which provides information on the proportion of data points below the fitted quantile values and bias in the estimates compared to original data.

B.5.3 Regression modelling

Generalised linear mixed modelling for the presence of pathogens given FIB or MST marker concentrations was conducted using the R package lme4 (Bates et al, 2015). The model used a binomial dependant variable with a logit linking function.

The univariate modelling approach assumed fixed effects of variables across all sites and random intercepts between sites to account for repeated measurements at each site. GLMM was chosen due to its ability to handle complex data structures and provide robust insights into the relationships between variables.

To avoid biasing the modelling outputs due to low numbers of samples at some sites, data was only considered from sites with 5 or more waters samples with no missing pathogen or indicator data.

Model fitting was conducted on the $\log_{10}$ transformed data for pathogen and MST marker concentrations, conductivity and turbidity measures.

The regression modelling of *Campylobacter* concentration dependant on multiple independent variables used the R package mgcv (Wood et al, 2017). As for univariate modelling, random intercepts were used to account for repeated measures at each site. Evaluation of candidate models followed the process suggested by Zurr et al (2009). As well as linear models, additive models were explored by allowing each independent variable in turn to be represented by thin plate regression splines via the mgcv function s().

APPENDIX C: SAMPLING SITES

Sites from which samples were collected during this project are listed in Table 58. During data analysis, sites were kept anonymous apart from an indication as to whether the site is in the North or South Island. This distinction was made to take into account general differences in temperatures and timing of farming activities such as lambing and calving, which can be a source of zoonotic pathogens in the environment.

Table 58: Location of sampling sites and observed dominant contamination sources

Site	Region	Observed local dominant contaminant source(s)	Phase	Comment
Fosters Bay unnamed stream	Auckland	Urban	3	Piped stream
Oakley Creek at waterfall	Auckland	Urban	3	
Oteha Stream at Days Bridge	Auckland	Urban	1,2,3	
Heathcote River at Ford Rd	Canterbury	Urban	3	
Heathcote River at Catherine St	Canterbury	Urban	1	Saline intrusion
Taylor River at Riverside Park	Marlborough	Urban	1,2	
Maitai River at Avon Terrace	Nelson	Urban	3	
Otepuni Creek at Nith Street	Southland	Urban	1,2,3	
Te Henui Stream mouth	Taranaki	Urban	3	Dairy upstream in catchment
Waitawhiriwhiri Stream at Edgecumbe Park	Waikato	Urban	3	
Porirua Stream at Town Centre	Wellington	Urban	1,2	
Sawyers Creek at Dixon Park	West Coast	Urban	1,2	
Pongakawa at State Highway 2	Bay of Plenty	Dairy	3	
Selwyn River at Coes Ford	Canterbury	Dairy	1,2,3	1998-2000 FMRP Study
Cam River upstream of Bramleys Road	Canterbury	Dairy	3	
Aurere River at Pekerau Rd	Northland	Dairy	3	
Waiotu River at State Highway 1	Northland	Dairy	1,2,3	

Site	Region	Observed local dominant contaminant source(s)	Phase	Comment
Moffat Creek at Moffat Road	Southland	Dairy	1,2,3	
Manganui River downstream of Everett Park	Taranaki	Dairy	2	Cultural significance
Manganui River at Everett Park	Taranaki	Dairy	3	Changed to council monitoring site
Pātea River at King Edward Park Stratford	Taranaki	Dairy	3	
Timaru Stream at Tataraimaka	Taranaki	Dairy	3	
Waitara River at Bertrand Rd	Taranaki	Dairy	1,2,3	
Waitara River downstream Manganui River	Taranaki	Dairy	2	Cultural significance
Waingongoro River at Ohawe Beach	Taranaki	Dairy	3	
Waingongoro River at Presbyterian camp Eltham	Taranaki	Dairy	3	
Motupipi downstream of Powell Creek	Tasman	Dairy	2	
Neds Creek at Murchison	Tasman	Dairy	2	
Powell Creek at Golden Bay	Tasman	Dairy	2	
Piako River at Paeroa-Tahuna Rd	Waikato	Dairy	1	
Waitoa River at Lansdowne Road bridge	Waikato	Dairy	3	
Kōpuaranga River at Stuarts	Wellington	Dairy	2	Cultural significance
Whangaehu River at Masterton Stronvar Rd	Wellington	Dairy	2	Cultural significance
Whangaehu River 250 m from Ruamahanga River	Wellington	Dairy	2	Cultural significance
Papakura River at Alfriston-Ardmore Rd	Auckland	Sheep and Beef	2,3	
Kaiate River at Kaiate Falls Rd	Bay of Plenty	Sheep and Beef	1,2,3	
Wharekōpae River at Rere Rockslide	Gisborne	Sheep and Beef	1,2	
Mangakuri River upstream of Mangakuri Road bridge	Hawkes Bay	Sheep and Beef	3	
Mangaorapa Stream at Mangaorapa Road	Hawkes Bay	Sheep and Beef	3	
Papakiri Sandy Creek upstream of Tutira Outlet	Hawkes Bay	Sheep and Beef	3	
Pōrangahau River at Kates Quarry	Hawkes Bay	Sheep and Beef	3	
Tukipo River at State Highway 50	Hawkes Bay	Sheep and Beef	3	
Waingongoro Stream at Waimārama Road	Hawkes Bay	Sheep and Beef	3	

Site	Region	Observed local dominant contaminant source(s)	Phase	Comment
Wairoa River at State Highway 2	Hawkes Bay	Sheep and Beef	2	
Manakau Stream at State Highway 1 bridge	Manawatū-Whanganui	Sheep and Beef	1,2	
Wakapuaka River at Hira	Nelson	Sheep and Beef	2	Cultural significance
Wakapuaka River at Māori Pa Rd	Nelson	Sheep and Beef	2	Cultural significance
Wakapuaka River at Paremata Flats Reserve	Nelson	Sheep and Beef	1,2,3	Cultural significance
Waipoua River at Akura Rd	Wellington	Sheep and Beef	2	Cultural significance
Kōpuaranga River Bluff at Rangitūmau Rd	Wellington	Sheep and Beef	2	Cultural significance
Ngongotahā Stream at Domain	Bay of Plenty	Dairy, Sheep and Beef	3	
Ōreti River at Iron Bridge	Southland	Dairy, Sheep and Beef	3	1998-2000 FMRP Study
Avon River at Kerrs Reach	Canterbury	Wild bird/ Urban	2,3	
Avon River at Antigua Boat Sheds	Canterbury	Wild bird/ Urban	2,3	1998-2000 FMRP Study
Lake Rotoroa at Café	Waikato	Wild bird/Urban	3	1998-2000 FMRP Study
Henley Lake at Masterton	Wellington	Wild bird/ Urban	2	
Ashburton River at State Highway 1	Canterbury	Wild bird/Sheep and Beef	2,3	
Ashburton River below Rail Bridge	Canterbury	Wild bird/Sheep and Beef	3	
Hātea River at Whangārei Falls	Northland	Wild bird/Sheep and Beef	1,2	1998-2000 FMRP Study, cultural significance
Hātea River upstream of Whangārei Falls	Northland	Wild bird/Sheep and Beef	3	Cultural significance, relocated upstream of drain
Lake Waiporohita	Northland	Wild bird/Dairy	3	
Kerikeri River near Stone Store	Northland	Wild bird/Dairy	3	
Waiwhakaiho River adjacent to Lake Rotomanu	Taranaki	Wild bird/Dairy	3	
Ōwairoa River downstream of Waitoa Rd	Bay of Plenty	Low impact	2	Relocated upstream of contamination
Ōwairoa River at Waitoa Rd	Bay of Plenty	Low impact	2, 3	
Wakapuaka River at Duckpond Rd	Nelson	Low impact	3	

Site	Region	Observed local dominant contaminant source(s)	Phase	Comment
Punaruku River at Russell Rd	Northland	Low impact	3	
Waipapa River in Waihou Valley	Northland	Low Impact	3	
Kauaeranga River at Smiths	Waikato	Low Impact	3	1998-2000 FMRP Study
Hutt River at Kaitoke	Wellington	Low Impact	2	
Arahura River at State Highway 6	West Coast	Low Impact	2	cultural significance

APPENDIX D: ENVIRONMENTAL DATA

Environmental factors affect the transport of contaminants to water bodies and the survival of micro-organisms in the water. The survival of microorganisms has been shown to be influenced by indigenous micro-organisms, temperature, and sunlight (Blaustein et al, 2013; Korajkic et al, 2013; Medema et al, 1997; Pachepsky et al, 2014; Yoneda et al, 2024). The measurement of rainfall, river flow and turbidity (water clarity) have each been proposed as predictors of the transportation of faecal contaminants to water bodies (McBride et al, 2002; United States Environmental Protection Agency, 2012; World Health Organization, 2021).

D.1 Field data collection

Physio-chemical field parameters that influence the transportation (rainfall and turbidity) and the survival of micro-organisms (pH, conductivity, dissolved oxygen, and water temperature) were measured. Not all councils had the equipment available for the measurements, so while water temperature was measured in almost all samples, turbidity was measured in only 49% (Table 59). Fourteen sites recorded turbidity as Formazin Nephelometric Units (FNU) for 10 or more samples. At three of these sites some data was reported as Nephelometric Turbidity Units (NTU).

Table 59: Physiochemical parameters measured in water samples

Parameter	Units	Percentage of samples measured	Median	Range	25th to 75th percentile
Water Temperature	°C	99%	14.4	5.4 - 26.8	12.2 - 16.9
pH		71%	7.54	4.63 - 9.91	7.15 - 7.93
Dissolved oxygen (DO)	mg/L	95%	11.6	0.2– 19.0	9.9 - 10.8
Conductivity	µS/cm	68%	180	2.74 – 48,337	117 - 259
Turbidity	NTU	10%	1.78	0 - 149	0.67 - 3.8
	FNU	40%	2.48	0 - 1929	0.93 - 7.26

In consultation with council hydrologists, rainfall and flow sites relevant to the sampling sites were identified. Not all sites had flow or rainfall routinely measured in locations appropriate to describing the rainfall effects or flow levels of the site. Hourly rainfall (mm) and average hourly flow rates were collected for the period of 3 days (72 hours) prior to water sampling.

Given the effects of rainfall and flow on indicator or pathogen levels are site specific, the national level QMRA model (sections 10 to 13) does not include rainfall or flow as variables. In future analysis of the data at a site level, rainfall and flow should be included in the analysis.

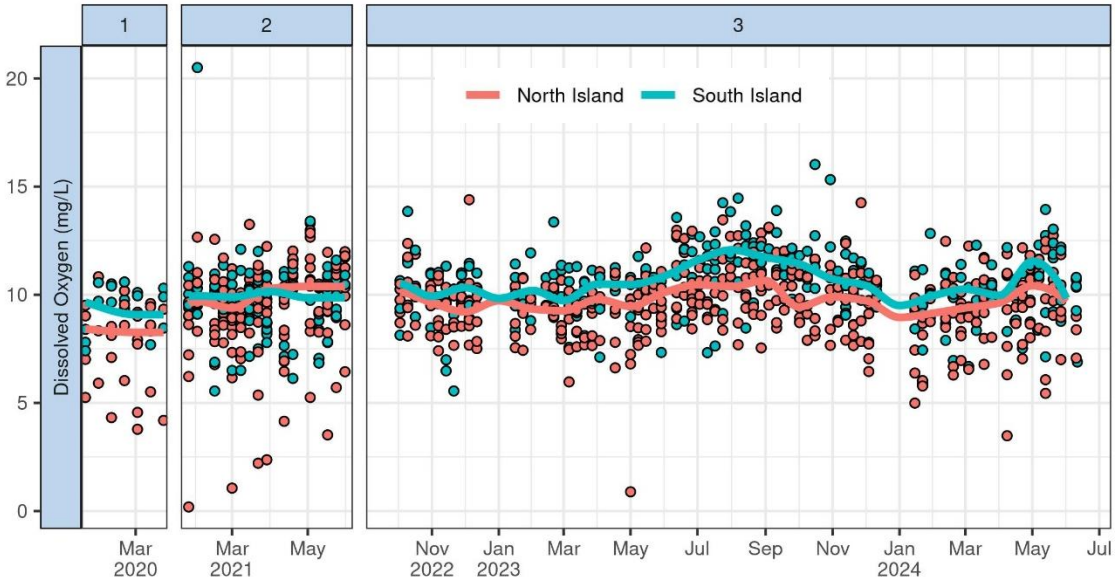
D.2 Conductivity, PH and dissolved oxygen

Conductivity was less than 1000 µS/cm for all but 21 samples, where higher levels indicated some saline intrusion (Table 59). These samples were not used for comparing associations between *E. coli* and pathogens.

Low pH (<6.50) occurred once at five sites and 45 times at four sites which indicated that low pH was characteristic of those sites. At 12 sites the pH was greater than 8.50 (38 samples). For one site this appeared to be characteristic of the site with high pH being recorded on 14 sampling occasions. On only two occasions was pH greater than 9.00, with the highest 9.91 being recorded in the presence of an algal bloom.

There were 38 samples from 13 sites with low DO (less than 6.5 mg/L). One site had DO below 4 mg/L on five occasions, and while DO increased in autumn all readings were less than 6.5, indicating it had ongoing low water quality. There was additional aeration at the site when the highest DO was recorded. The DO median by week for South Island samples and North Island samples showed that the water bodies in the South Island had comparatively higher median DO than the North Island (Figure 39).

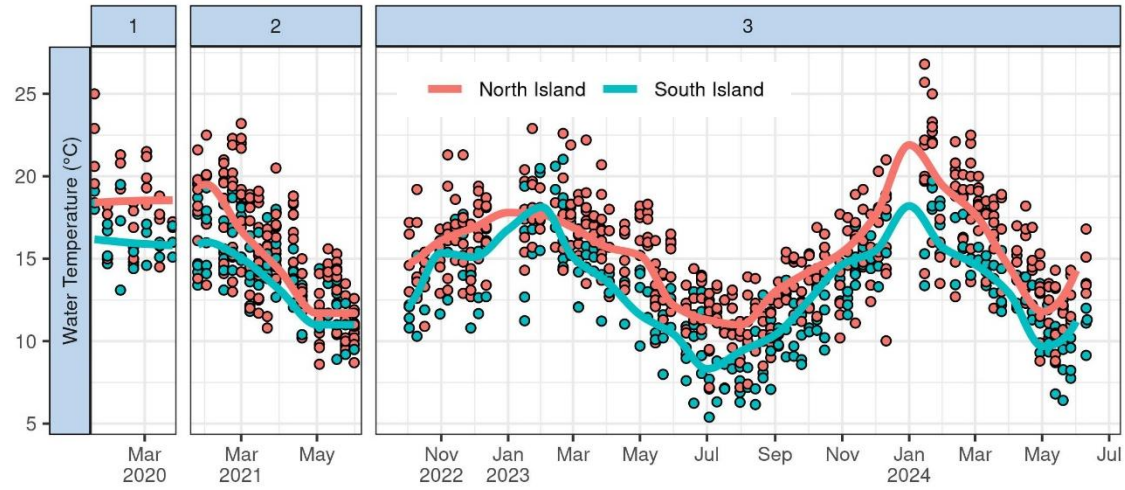
Figure 39: Dissolved oxygen data recorded during the three phases of the study, coloured by North or South Island with time phased weekly medians by island.



D.3 WATER TEMPERATURE AND SEASON

Water temperature ranged from 5.4 – 26.8°C and followed a seasonal pattern (Figure 40). There were 66 samples when the water body temperature was <10°C (April to October), 469 samples where temperatures were between 10°C and <15°C, 337 samples between 15 and <20°C and 46 samples taken when the water temperature was at, or above, 20°C (December to March) with two of those samples being above 25°C in January).

Figure 40: Water temperature data from the three phases of the study, coloured by North and South Island sites and showing time phased weekly medians by island.



APPENDIX E: DOSE RESPONSE

E.1 **Campylobacteriosis**

The simulation model uses the dose-response models for campylobacteriosis (see Table 46) based on the work of (Medema et al, 1996) for the probability of infection and data from (Black et al, 1988) and unpublished data from (Teunis et al, 2018) for doses of the order of 10^4 or less, for the probability of illness given infection. The maximum simulated dose generated by the model is of the order of 10^4.

There has been no subsequent feeding trial data at these low doses and there is large uncertainty in doses consumed from outbreaks. For this reason, the Medema model used for the previous freshwater QMRA will be used as the primary dose-response relationship for this project. Providing continuity of approach until new information becomes available to confirm or update the dose-response relationship.

An alternative fixed parameter dose-response model has been developed using data from feeding trails (Teunis et al, 2018) and used in a number of overseas QMRA models (United States Environmental Protection Agency, 2024):

$$P_{\text{inf}\,|\,D} = 1 - \ {}_1F_1(\alpha, \alpha+\beta, -D) \quad \text{Feeding trials:} \quad \alpha = 0.44, \quad \beta = 0.51 \qquad \textbf{(B.1)}$$

$$P_{\text{ill}\,|\,inf,D} = 1 - \left(1 + \frac{D}{\eta}\right)^{-r} \quad \text{Feeding trials:} \quad \eta = 0.88, \quad r = 0.06 \qquad \textbf{(B.2)}$$

In the above model the probability of illness given infection, is dependent on dose. The point estimates of the dose-response parameters in equations B.1 and B.2 are the median values of the parameter estimates generated from Markov Chain Monte Carlo analyses.

The Markov Chain Monte Carlo analyses also produced bivariate models for dose-response parameters which further incorporate uncertainty in the dose-response relationship for infection or illness. In the simulation model, the pairs of parameters of equations B.1 and B.2 are sampled from a transformed bivariate distribution:

$$z = \log(\alpha+\beta) \qquad \text{mean}(z) = 0.054 \qquad \text{var}(z) = 1.070$$

$$w = \log\left(\frac{\frac{\alpha}{\alpha+\beta}}{1-\frac{\alpha}{\alpha+\beta}}\right) \qquad \text{mean}(w) = -0.177 \qquad \text{var}(w) = 1.303 \qquad \textbf{(B.3)}$$

$$\text{cov}(w,z) = -0.041$$

and

$$z = \log(r+\eta) \qquad \text{mean}(z) = 0.00489 \qquad \text{var}(z) = 0.993$$

$$w = \log\left(\frac{\frac{r}{r+\eta}}{1-\frac{r}{r+\eta}}\right) \qquad \text{mean}(w) = -2.744 \qquad \text{var}(w) = 1.337 \qquad \textbf{(B.4)}$$

$$\text{cov}(w,z) = 0.010$$

The simulation model has been run in a two-dimensional framework using the R mc2d package (Pouillot and Delignette-Muller, 2010). The uncertainty in the dose-response parameters as described by the above bivariate distribution has been implemented with 500 iterations in the uncertainty dimension and 50,000 iterations in the variability dimension (as used in the base model).

Using the (Teunis et al, 2018) dose-response in the simulation model, results in infection rate estimates that are 5 to 9 times higher than estimated by the (Medema et al, 1996) dose-response model (Table 60). The largest difference being for water samples with the lowest concentration of *E. coli*. However, the Teunis model probability of illness given infection has lower values than modelled in the base model, making the final rate of illness estimated by the Teunis model to be approximately twice that of the base model using the Medema dose response.

Table 60: Comparison of simulation outputs for three different campylobacteriosis dose-response models, using data collected over the full year.

E. coli /100 ml	Percentage of swims resulting in infection or illness from *Campylobacter*[a]					
	Base Model: Medema et al, 1996 and Uniform (0.2 to 0.6)		Teunis et al, 2018 (Feeding trials, fixed parameters)		Teunis et al, 2018 (Feeding trials, bivariate normal parameters)	
	Infection	Illness	Infection	Illness	Infection	Illness
100	2.5	1.0	22.5	2.5	22.0 (5.5 – 32.5)[b]	2.0 (<0.5 – 13.5)[b]
300	6.0	2.5	37.0	5.0	35.5 (10.5 – 49.5)	4.5 (0.5 – 24.5)
500	8.0	3.0	42.0	6.5	41.5 (13.5 – 55.5)	6.0 (0.5 – 30.5)
800	9.5	4.0	46.5	7.5	45.0 (15.5 – 58.5)	6.5 (0.5 – 34.0)
1000	9.5	4.0	46.0	7.5	45.5 (16.0 – 59.5)	7.0 (0.5 – 35.0)
> 2000	10.0	4.0	48.0	8.0	47.0 (17.5 – 61.0)	7.5 (0.5 – 36.5)

Notes:
a: Percentage of swimming events estimated to result in infection or illness rounded to nearest multiple of 0.5.
b: Median 50th (2.5th – 97.5th) Hazen percentiles of estimates due to uncertainty in the dose-response parameter values.

Incorporating the uncertainty in the dose response parameters, as modelled by Teunis et al. (2018), results in large intervals of the infection and illness estimates (right hand columns of Table 60). These uncertainty intervals reflect the limited and variable data the Teunis models are based on. Further data is required to improve our understanding of the probability of infection and illness due to campylobacteriosis.

E.2 **Cryptosporidiosis**

Two dose-response models have been considered:

i. Exponential model (Equation B.5) as provided in the (United States Environmental Protection Agency, 2024) technical support document for alternative recreational criteria, as used for the base simulation model.

$$P_{\inf|D} = 1 - exp(-rD) \quad r = 0.09 \qquad \textbf{(B.5)}$$

ii. Fractional Poisson model (Equation B.6) as presented in (M. Messner and P. Berger, 2016).

$$P_{\inf|D} = P\left(1 - exp(-D)\right) \quad P = 0.739 \qquad \textbf{(B.6)}$$

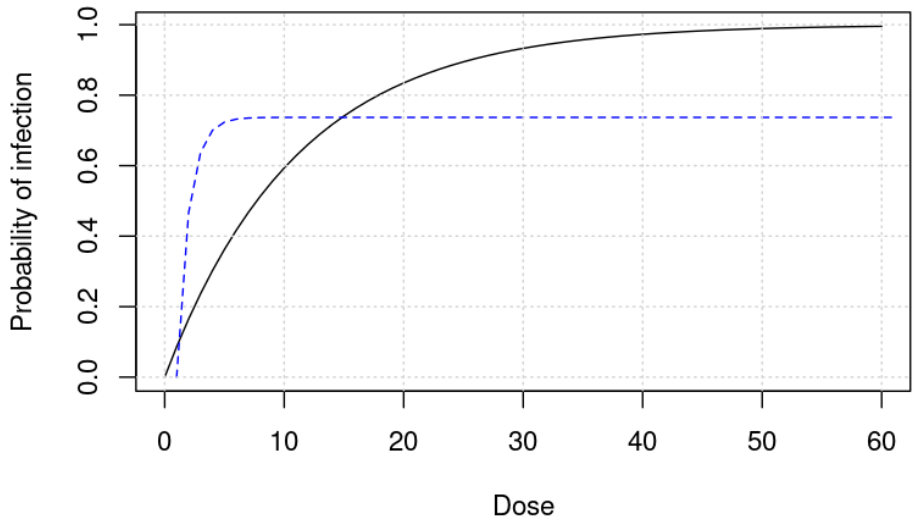
The exponential model assumes that everyone is susceptible to infection given a sufficient dose *D*, and the *r* parameter is a measure of the probability that oocyst will survive and cause an infection.

When the *r* value is set to one as in the fractional Poisson model, it is assumed a dose of a single oocyst will cause infection. However, the fractional Poisson model, assumes there is a proportion of the population *P* which are immune to infection.

The fractional Poisson model (B.6) has a rapid increase of probability of infection between doses of 0 and 5 oocysts, followed by no further increase in probability of infection beyond 0.74 (Figure 41). In contrast the exponential model (B.5) has a lower probability of infection at lower doses, but infection rates continue to increase with dose until a dose of 60 oocysts, beyond which, the model estimates everyone would get infected.

Both these models are based on the same feeding trials, which used healthy volunteers, and detection of oocysts in the volunteer's stools. Of the 119 subjects, only 8 were exposed to the lowest intended dose of 10 oocysts. Ten oocysts is higher than most of the doses estimated by the simulation model. Therefore, the simulation doses are in the zone of greatest uncertainty in the dose response, but the two approaches can be interpreted to inform an uncertainty envelope for the simulation estimates.

Figure 41: Exponential (solid line) and Fractional Poisson dose-response models for cryptosporidiosis.



The results of running the simulation model using the two dose response relationships are given in Table 61. The estimated rates of illness per 1000 swim events using the fractional Poisson model is approximately five times the rates for the exponential model.

Table 61: Comparison of simulation outputs using enterococci as the FIB for two different cryptosporidiosis dose-response models.

Enterococci (MPN/100 ml)	Percentage of swims resulting in infection or illness			
	Base Model: Exponential and Uniform (0.3 to 0.7)		Fractional Poisson Model and Uniform (0.3 to 0.7)	
	Infection	Illness	Infection	Illness
100	0.6	0.3	2.7	1.4
1,000	0.8	0.4	3.9	2.0
10,000	1.1	0.6	5.4	2.8

E.3 **Giardiasis**

Two dose-response relationships for *Giardia duodenalis* have been implemented in the simulation model, based on the exponential dose-response relationship (Equation B.5).

The dose-response relationship used in the base model has the fraction of cysts which survive ingestion and cause infection as $r = 0.0199$ (Rose et al, 1991). This parameter estimate is derived from the data from two feeding studies.

The first feeding study involved feeding known doses of *Giardia* cysts in a gelatine capsule to male volunteers aged 21 to 56 years old (Rendtorff, 1954). Doses ranged from 1 to 10^6 cysts. While infection was indicated by presence of giardia in faecal stools, the definition of illness is not clear from the journal paper. None of the controls or 5 males who consumed a single cyst became infected. Eight out of 22 men with a dose of 10-25 cysts, and all thirteen men given doses between 10^2 and 10^6 cysts became infected. In a second trial, one of two men given 10^2 cysts in water became infected (Rendtorff and Holt, 1954).

The probability of illness given infection is assumed to be independent of dose and modelled by a uniform distribution with a range of 0.3 to 0.7 (Jahne et al, 2024).

The second dose-response considered an exponential relationship for the probability of infection given dose, and a point estimate (m) of probability of illness given infection, but with uncertain parameters (T. Burch, 2020). Parameters were fitted to a Bayesian random-effects model, taking into account variability and uncertainty in, and censoring of, the available source data. This approach was implemented using 2-dimensional Monte Carlo simulations using the R package mc2d (Pouillot and Delignette-Muller, 2010).

The dose-response parameters are modelled by:

$$logit_{10}(r) = logit_{10}(r_0) + Normal(0, \sigma_r) r_0 = 0.016, \quad \sigma_r = 0.52$$

$$logit_{10}(m) = logit_{10}(m_0) + Normal(0, \sigma_m) \quad m_0 = 0.38, \quad \sigma_m = 0.93$$

The source data for the modelling were from two Rendtorff experimental feeding studies and three waterborne outbreaks.

The three waterborne outbreaks (Kent et al, 1988; Kirner et al, 1978; Nygård et al, 2006) recorded numbers of reported illness (3800, 240 and 2500), however there is uncertainty and variability in the dose the outbreak population would have ingested. As well as uncertainty in the number of the exposed population who became infected, and possible over-reporting of illness associated with *Giardia* during an outbreak.

The dose-response approaches used in the QMRA assume there is no aggregation of cysts.

Table 62 compares the simulation model outputs for the two dose response approaches. The intervals for the Burch (2020) dose-response model are the uncertainty intervals generated by the simulation model. The uncertainty in the dose-response model suggests a higher rate of illness is possible than predicted by the point estimate base model.

Table 62: Comparison of simulation outputs using enterococci as the FIB for two different giardiasis dose-response models.

Enterococci (MPN/100 ml)	Percentage of swims resulting in infection or illness			
	Base Model: (Rose and Gerba, 1991)		2d MC model - parameters (Burch, 2020)	
	Infection	Illness	Infection	Illness
100	0.2	0.1	0.2 (< 0.1 – 1.9)	< 0.1 (< 0.1 – 0.9)
1000	0.3	0.1	0.3 (< 0.1 – 3.0)	0.1 (< 0.1 – 1.4)
10,000	0.5	0.2	0.5 (< 0.1 – 5.0)	0.1 (< 0.1 – 2.4)

APPENDIX F: SENSITIVITY ANALYSIS

F.1 Introduction

This appendix describes analyses to explore the sensitivity of the simulation model outputs to the model inputs and approach taken. The base model for comparison is the model used for the results presented in this report (section 13.1).

F.2 Simulation iterations and reporting accuracy

The model was run for 1,000 to 200,000 iterations, and the outputs were found to be stable to within 0.22% or 2.2 illnesses per 1,000 swim events for 50,000 or more iterations. Hence 50,000 iterations have been used for the results presented in this report and the percentage of swims results are reported rounded to multiples of 0.5.

F.3 Water ingestion rate

The sensitivity of the model outputs for campylobacteriosis infection to the amount of water ingested has been evaluated by two approaches:

- Using fixed volumes of ingested water.
- Setting the duration of swimming to be inversely related to the water ingestion rate of the swimmer.

F.3.1 *Variable verses fixed ingested water volumes*

The base model samples from a set of variable water ingestion values, ranging from 0 ml to 1230 ml of water (Figure 30). Previous international QMRA simulations (World Health Organization, 2016) have used a single fixed value for the volume of water ingested, a uniform distribution between bounds or a distribution of values based on study of DeFlorio-Barker et al. (2018) and Dufour et al (2017). To examine how the representation of water volume ingestion impacts the results, the simulation base model has been run with selected fixed values of water ingestion between 20 and 200 ml (Table 63).

Table 63: Simulation model estimation of Campylobacter infection given swimmers ingesting selected volumes of water (rounded to 0.5%). Highlighted columns are for the base model with variable ingestion volume and the closest model results using a fixed volume ingested.

E. coli range (MPN/100 ml)	Percentage of swims resulting in *Campylobacter* infection given people ingest the volume of water					
	Variable ingestion volume (6-12 year olds)	20 ml	40 ml	50 ml	100 ml	200 ml
100	2.5	1.5	2.0	2.5	4.0	6.0
300	6.0	3.5	5.0	6.0	8.5	12.0
500	8.0	5.0	7.5	8.0	11.0	15.0
800	9.5	6.5	9.0	10.0	13.0	17.0
1,000	9.5	6.5	9.0	10.0	13.0	17.0
> 2,000	10.0	7.0	9.5	11.0	14.0	18.0

The results in Table 63 show that model is sensitive to the choice of volume of water sampled, which would be expected given the dose is directly related to the volume of water ingested. A fixed volume of ingested water of 40-50 ml results in similar estimates of swimmer infection rate as produced by the model with the variable ingestion volume (highlighted columns in Table 63).

F.3.2 *Independence of water ingestion rate and duration of swimming activity*

In the base model, the duration of swimming is assumed to be independent of the rate that water is ingested. The highest volumes of water ingested will relate to the highest durations and highest ingestion rates.

If instead, it could be assumed that a child ingesting higher volumes of water is likely to tire more quickly and therefore swim for a shorter period of time.

To investigate the effect of this dependency on the simulation results, the 50,000 simulation swimming durations of the base model were adjusted:

$$\begin{aligned} duration_{adj} &= duration && ingestion\,rate \leq I \\ duration_{adj} &= duration\left(1-\gamma\frac{ingestion\,rate}{\max ingestion\,rate}\right) && ingestion\,rate > I \end{aligned} \quad \textbf{(G.1)}$$

where *I* is the ingestion rate below which there is no dependence on duration of swimming, and γ is a value between 0 and 1 which reduces duration compared to the initial simulation duration. Simulations were run for ingestion rate ***I*** of 50 and 100 ml/h and a reduction factor ***γ*** of 0.5 and 0.9 (Figure 42 and Table 64).

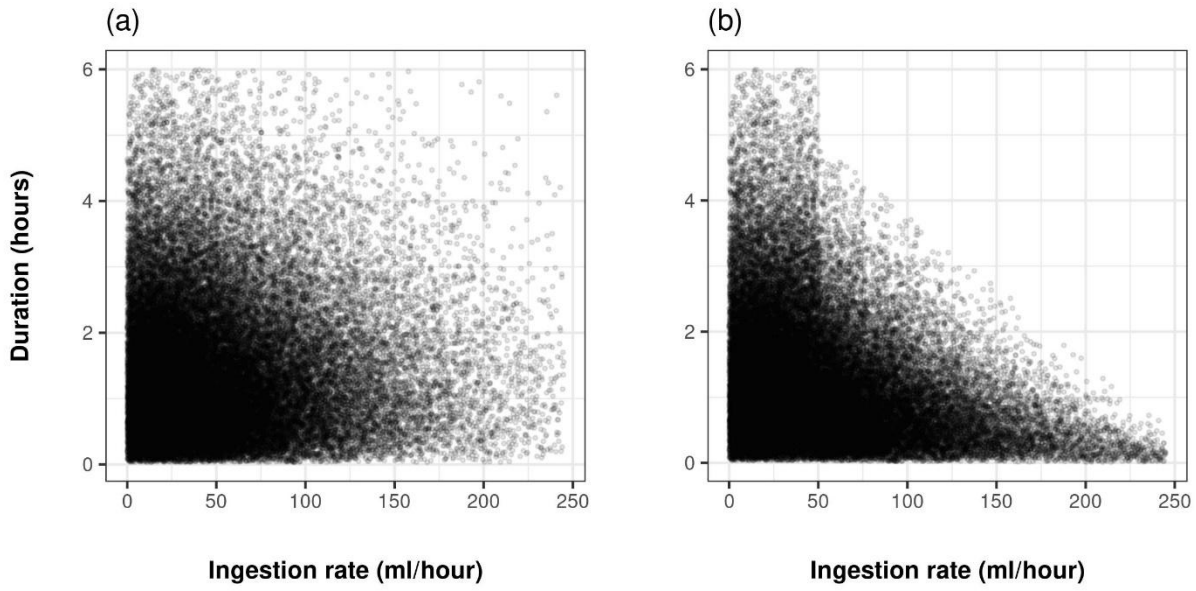
Table 64: Simulation model estimation of campylobacteriosis infection given swimmers ingesting volumes of water with different dependencies between the water ingestion rate and duration of the swim event.

E. coli range (MPN/100 ml)	Percentage of swims resulting in *Campylobacter* infection			
	Variable ingestion volume (6-12 year olds)	I = 100 ml/h γ = 0.5	I = 100 ml/h γ = 0.9	I = 50 ml/h γ = 0.9
100	2.7	2.6	2.5	2.3
300	6.0	5.9	5.7	5.3
500	8.1	7.9	7.5	7.1
800	9.7	9.4	9.2	8.8
1,000	9.7	9.5	9.4	9.0
> 2,000	10.2	10.1	9.8	9.3

In the base model, approximately 40% of swim events have an ingestion rate of over 50 ml/h and 15% of swim events have an ingestion rate of over 100 ml/h. The maximum adjustment for swim duration (I = 50 mL/h, γ = 0.9) resulted in a decrease of between 0.4 and 0.9 in the percentage campylobacteriosis infections (Table 64).

Given there is no data to support modelling this aspect of water intake and the modest conservative impact of the independence assumption on the results; the assumption of independence of water ingestion rate with swim duration is used in the simulation model.

Figure 42: Simulation swim event duration and water ingestion rate: (a) assuming two variables are independent and (b) assuming duration is likely to decrease with increasing ingestion of water for ingestion rate over 50 ml/h and a reduction factor of 0.9.



F.4 Analysis bias in collection of protozoa data in low volumes of water

The amount of water filtered for quantification of Cryptosporidium or Giardia varied between the phases of the project, but the target for all phases was at least 10 L. In some cases, less than 10 L was collected and in 37 cases less than 5 litres was collected. The reason for the smaller volumes of water was either equipment failure or the particles in the water clogged up the filter before the target water volume could be filtered.

For a time during the project water volumes less than 5 L were not analysed (21 water samples), due to concerns over measurement uncertainty in lower volumes. As protozoa can be associated with particles and sediments that may be present in streams (Atherholt et al, 1998), it is possible more higher protozoa concentrations would have been identified, if these samples had been analysed, and therefore a higher risk of infection and illness predicted. Two of the four highest recorded *Giardia* concentrations (42 and 25 cysts/10 L) were associated with filtered water volumes of 5 L or less and the three highest *Cryptosporidium* concentrations (140 to 21 oocysts/10L) where associated with filtered volumes less than 5 L.

To investigate the potential effect of not analysing these water samples for *Cryptosporidium*, the missing data in 21 data samples was replaced using data from the other 16 water samples with volumes less than 5 L which were analysed. The 5 highest concentration values out of the 16 data points were reused in the imputation of the 21 missing values.

There is a slight increase in the estimated infection and illness percentages, when missing data is imputed with comparable concentrations based on the volume of water collected and (Table 65). However, the increase is within the modelling uncertainty of 0.5% and the percentage of the infections and illness remains below the percentages estimated for campylobacteriosis.

Table 65: Comparison of simulation outputs using enterococci as the FIB for *Cryptosporidium* infection and illness with and without imputation of missing *Cryptosporidium* concentrations at low volumes of filtered water.

Enterococci (MPN/100 ml)	Percentage of swims resulting in infection or illness			
	Data as collected in the study		Imputed *Cryptosporidium* concentrations for missing data for water volumes less than 5 L	
	Infection	Illness	Infection	Illness
100	0.6	0.3	0.7	0.3
1,000	0.8	0.4	1.0	0.5
10,000	1.1	0.6	1.5	0.8

Using the Fractional Poisson (appendix E.2) instead of the exponential dose-response model used for Table 65 estimated an increase from two percent of swims resulting in illness to 2.3% when imputing for the missing values and an enterococci concentration of 1,000 MPN/100 ml.

APPENDIX G: GLOSSARY

25th percentile	The lowest value that is greater than 25% of all data points (also known as the first quartile (Q1) or lower quartile)
50th percentile	The value at which 50% of data points are below and 50% are above (also known as the second quartile (Q2) or median)
75th percentile	The value at which 75% of all data points fall below (also known as the third quartile (Q3) or upper quartile)
95th percentile	The value at which 95% of all data points fall below
Acute gastrointestinal illness	Sudden onset of gastrointestinal inflammation commonly leading to diarrhoea and/or vomiting
Adenovirus	A group of viruses first identified in human adenoid tissue which commonly cause respiratory symptoms
Allele	One of two or more variant sequences found at a specific genomic location (locus)
Asymptomatic	When a person affected by a particular medical condition displays no symptoms
Bacteriophage	A virus which infects bacteria
Campylobacter	A genus of pathogenic bacteria known to be a leading cause of food poisoning
Campylobacter coli	A species of thermophilic *Campylobacter* which is found in humans and animals
Campylobacter jejuni	The most common thermophilic *Campylobacter* species to be found in both human clinical samples and animals
Campylobacter lari	
Campylobacteriosis	Gastrointestinal disease caused by bacteria of the *Campylobacter* genus, a common cause of food poisoning
Case-control study	An observational study that compares two groups – those with a given disease or condition and those without to identify potential risk factors
Core genome multilocus sequence typing	A high-resolution molecular typing method which uses whole genome sequencing data allowing a much larger set of core genes to be compared than traditional MLST analysis
Enterococci	A genus of bacteria which naturally colonise the gut of humans and animals and as such can be used as an indicator of faecal contamination
Enumerate	To determine the number of
Escherichia coli	A group of bacteria which usually live in the gut of humans and animals without causing disease
Exposure	Contact with an infectious organism (e.g., via ingestion, inhalation, dermal contact). May be short-term (acute) or long-term (chronic)
Faecal coliforms	Bacteria that normally live in the gut of warm-blooded animals and as such are present in faeces

Faecal indicator bacteria	Bacteria found in the gut of warm-blooded animals whose presence in water can be used as an indicator of faecal contamination
Hazard characterisation	The second step in a QMRA in which the potential adverse health effects associated with the microbe(s) of concern are identified and described
Hazard identification	The first step in a QMRA in which the microorganisms of most concern for the study are identified together with the diseases or illnesses that they cause
Human health risk assessments	A method for assessing potential impacts on the health of a person, or population, from a given hazard
Land use category	Classifications of how a given land area is being utilised (e.g., residential, agricultural, commercial)
Limit of quantification	The lowest concentration of analyte which can be accurately measured by a given method
Maximum quantifiable concentration	The highest amount of a given analyte which can be accurately and reliably determined using a given method
Median	The middle number in an ordered numeric dataset (or average of the two middle numbers when there are an even number of data points)
Microbial water quality	A measure of the presence and concentration of microbial contaminants which may pose a health risk in a given body of water
Minimum quantifiable concentration	The lowest concentration which can be reliably and accurately determined using a given method
Monte Carlo simulation models	Models used to predict the probability of different outcomes of uncertain events
Microbial source tracking marker	A host-specific genetic marker used to differentiate between faeces from different species
Multilocus sequence typing	A molecular method used to characterise isolates of a bacterial species based on the sequences of internal fragments from multiple housekeeping genes
Multiplex PCR	A PCR technique in which multiple DNA or RNA targets are amplified simultaneously in a single reaction
Campylobacteriosis	Gastrointestinal disease caused by bacteria of the *Campylobacter* genus, a common cause of food poisoning
Case-control study	An observational study that compares two groups – those with a given disease or condition and those without to identify potential risk factors
Core genome multilocus sequence typing	A high-resolution molecular typing method which uses whole genome sequencing data allowing a much larger set of core genes to be compared than traditional MLST analysis
Enterococci	A genus of bacteria which naturally colonise the gut of humans and animals and as such can be used as an indicator of faecal contamination
Enumerate	To determine the number of

Escherichia coli	A group of bacteria which usually live in the gut of humans and animals without causing disease
Exposure	Contact with an infectious organism (e.g., via ingestion, inhalation, dermal contact). May be short-term (acute) or long-term (chronic)
Faecal coliforms	Bacteria that normally live in the gut of warm-blooded animals and as such are present in faeces
Faecal indicator bacteria	Bacteria found in the gut of warm-blooded animals whose presence in water can be used as an indicator of faecal contamination
Hazard characterisation	The second step in a QMRA in which the potential adverse health effects associated with the microbe(s) of concern are identified and described
Hazard identification	The first step in a QMRA in which the microorganisms of most concern for the study are identified together with the diseases or illnesses that they cause
Human health risk assessments	A method for assessing potential impacts on the health of a person, or population, from a given hazard
Land use category	Classifications of how a given land area is being utilised (e.g., residential, agricultural, commercial)
Limit of quantification	The lowest concentration of analyte which can be accurately measured by a given method
Maximum quantifiable concentration	The highest amount of a given analyte which can be accurately and reliably determined using a given method
Median	The middle number in an ordered numeric dataset (or average of the two middle numbers when there are an even number of data points)
Microbial water quality	A measure of the presence and concentration of microbial contaminants which may pose a health risk in a given body of water
Minimum quantifiable concentration	The lowest concentration which can be reliably and accurately determined using a given method
Monte Carlo simulation models	Models used to predict the probability of different outcomes of uncertain events
Microbial source tracking marker	A host-specific genetic marker used to differentiate between faeces from different species
Multilocus sequence typing	A molecular method used to characterise isolates of a bacterial species based on the sequences of internal fragments from multiple housekeeping genes
Multiplex PCR	A PCR technique in which multiple DNA or RNA targets are amplified simultaneously in a single reaction
Negative predictive value (NPV)	Probability that a sample doesn't have pathogen
Norovirus	A highly contagious virus which is the most common cause of viral gastroenteritis
Notifiable disease	Infections or diseases which must be reported

Oocysts	Environmentally resistant, infectious propagules of parasitic protozoa
Pathogen	An organism which causes disease in another host organism
Pathogenic	The ability of a microorganism to cause disease
Polymerase chain reaction	A molecular method used to rapidly produce large amounts of a target DNA sequence from a small starting amount of DNA sample
Positive predictive value (PPV)	Probability that a sample actually has pathogen
Prevalence	Occurrence of an organism within a certain area or sample; or occurrence of an illness or disease within a defined population
Protozoa	A diverse group of single celled eukaryotic microorganisms which includes the pathogens *Giardia* and *Cryptosporidium*
Quantitative microbial risk assessment	A tool for estimating human health risks associated with the presence of pathogens in food or the environment
Quantitative PCR	A method used to simultaneously amplify and quantify a target DNA molecule
Qualitative data	Non-numerical information (e.g., descriptors, observations
Quantitative data	Any information which can be quantified, counted or measured
Salmonellosis	A gastrointestinal bacterial infection caused by members of the Salmonella genus (e.g., *S. enterica*, *S. Enteritidis*, *S. Typhimurium*) and commonly associated with food poisoning
Sanitary survey	An on-site assessment of potential routes for faecal contamination
Sensitivity	Sensitivity – ability of a test to correctly identify the presence of pathogen (true positive rate)
Sequelae	Complications arising from a previous disease, infection or injury
Serotype	A distinct subgroup within a species that have the same number and type of surface antigens
Shiga toxin	A potent biological poison produced by specific bacterial pathogens including *Shigella dysenteriae* and Shiga-toxin producing *E. coli* (STEC)
Specificity	Specificity – ability of a test to correctly identify samples without pathogen (true negative rate)
Shiga toxin-producing *E. coli*	A pathogenic group of *E. coli* which produce one or more extracellular shiga toxins and are capable of causing severe gastrointestinal infection
Strain	A genetic variant or subtype within a species
Surface water	Water located above ground and contained within a body such as a lake or river
Thermotolerant	The ability of an organism to survive and grow at high temperatures
Total coliforms	A group of bacteria commonly found in the intestines of mammals and as such their presence in water is often used as an indicator of potential faecal contamination

Total thermophilic *Campylobacter*	Determined from standard *Campylobacter* culture test before confirming for the presence of *Campylobacter jejuni* and *Campylobacter coli*. The concentration may be greater than the sum of *C. jejuni* and *C. coli* due to the presence of other less common thermophilic *Campylobacter* species
Virulence	A measure of a pathogens ability to cause damage to a host, with more virulent pathogens causing more severe disease
Whole genome sequencing	A method for determination of the entire genomic sequence of an organism, including both coding (genes) and non-coding regions
Wild bird	Birds that live in the wild, particularly around water sources e.g., ducks, geese
Zoonotic pathogen	Infectious microorganisms that can be transmitted from animals to humans

APPENDIX H: ABBREVIATIONS

AEEC	Atypical enteropathogenic *E. coli*
AGI	Acute gastrointestinal illness
cgMLST	Core genome multilocus sequence typing
CT	Cycle threshold
D	Dose
DAEC	Diffusely adherent *E. coli*
DALYs	Disability-adjusted life years
DBLQ	Detected below the limit of quantification
dl	Detection limit
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
EaggEC	Enteroaggregative *E. coli*
eDNA	Environmental DNA
EHEC	Enterohaemorrhagic *E. coli*
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ESR	Institute of Environmental Science and Research Ltd
ESRD	End stage renal disease
ETEC	Enterotoxigenic *E. coli*
EU	European Union
FIB	Faecal indicator bacteria
FMRP	Freshwater Microbiology Research Programme
FNU	Formazin nephelometric units
GBS	Guillain-Barré syndrome
GC	Gene copies
GIS	Geographic information systems
HAdV	Human adenovirus
HUS	Haemolytic uraemic syndrome
IBD	Irritable bowel disease
IBS	Irritable bowel syndrome
LINZ	Land Information New Zealand
LOD	Limit of detection
LOQ	Limit of quantification
MfE	Ministry for the Environment
MLST	Multilocus sequence typing
MoH	Ministry of Health
MPN	Most probable number
MST	Microbial source tracking
ND	Not detected
NIWA	National Institute of Water and Atmospheric Research
NOF	National Objective Framework
NoV	Norovirus
NPS-FM	National Policy Statement Freshwater Management
NPV	Negative predictive value

NTU	Nephelometric turbidity units
PCR	Polymerase chain reaction
PHF Science	New Zealand Institute for Public Health & Forensic Science (formerly ESR)
PI-IBS	Post-infectious irritable bowel syndrome
PPMCC	Pearson's product movement correlation coefficient
PPV	Positive predictive value
QMRA	Quantitative microbial risk assessment
qPCR	Quantitative polymerase chain reaction
Rd	Road
ReA	Reactive arthritis
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
RWQC	Recreational water quality criteria
sp.	Species (singular)
spp.	Species (plural)
ST	Sequence type
STEC	Shiga toxin-producing *E. coli*
subsp.	Subspecies
TTC	Total thermophilic *Campylobacter*
TTP	Thrombotic thrombocytopenic purpura
USEPA	United States Environmental Protection Agency
VTEC	Verocytotoxin toxin-producing *E. coli*
wgMLST	Whole genome multilocus sequence typing
WGS	Whole genome sequencing
WHO	World Health Organization
YLL	Years of life lost

REFERENCES

Adlam S, Perera, S, Lake, R, Campbell, D, Williman, J, Baker, M. (2011). Acute gastrointestinal illness in New Zealand: a community study. *Epidemiology and Infection*, 139, 302-308.

Ahmed W, Harwood, V, Nguyen, K, Young, S, Hamilton, K, Toze, S. (2016). Utility of *Helicobacter* spp. associated GFD markers for detecting avian fecal pollution in natural waters of two continents. *Water Research*, 88, 613-622.

Ahmed W, Kirs, M, Gilpin, B. (2011). Source tracking in Australia and New Zealand: Case studies. In C Hagedorn, AR Blanch, and VJ Harwood (Eds.), *Microbial source tracking: methods, applications, and case studies* (pp. 485-513). Springer New York. https://doi.org/10.1007/978-1-4419-9386-1_21.

Ahmed W, Korajkic, A, Gabrewold, M, Payyappat, S, Cassidy, M, Harrison, N, Besley, C. (2024). Assessing the nucleic acid decay of human wastewater markers and enteric viruses in estuarine waters in Sydney, Australia. *Science of The Total Environment*, 926, 171389.

Ahmed W, Payyappat, S, Cassidy, M, Besley, C. (2019). A duplex PCR assay for the simultaneous quantification of Bacteroides HF183 and crAssphage CPQ_056 marker genes in untreated sewage and stormwater. *Environment International*, 126, 252-259.

Ahmed W, Schoen, ME, Soller, J, Harrison, JC, Hamilton, KA, Gebrwold, M, Simpson, SL, Payyappat, S, Cassidy, M, Harrison, N, Besley, C. (2024). Site-specific risk-based threshold (RBT) concentrations for sewage-associated markers in estuarine swimming waters. *Sci Total Environ*, 929, 172448.

Ajene A, Fischer Walker, C, Black, R. (2013). Enteric pathogens and reactive arthritis: a systematic review of *Campylobacter*, *salmonella* and *Shigella*-associated reactive arthritis. *Journal of Health, Population and Nutrition*, 31, 299-307.

APHA. (2017a). 9223 Enzyme substrate coliform test. 9223 B enzyme substrate test. In *Standard methods for the examination of water and wastewater*. American Public Health Association, American Water Works Association, Water Environment Federation.

APHA. (2017b). 9230 Fecal Enterococcus/Streptococcus groups. 9230 D fluorogenic substrate Enterococcus test. In *Standard methods for the examination of water and wastewater*. American Public Health Association, American Water Works Association, Water Environment Federation.

APHA. (2017c). 9260 Detection of pathogenic bacteria 9260 G *Campylobacter*. In *Standard methods for the examination of water and wastewater*. American Public Health Association, American Water Works Association, Water Environment Federation.

APHA. (2017d). 9260 Detection of pathogenic bacteria. 9260 B – *Salmonella*. In *Standard methods for the examination of water and wastewater*. American Public Health Association, American Water Works Association, Water Environment Federation.

Armstrong B, Pattis, I, Horn, B, Lopez, L, Cressey, P, Soboleva, T. (2024). *Annual report concerning foodborne diseases in New Zealand 2023.* ESR client report for Ministry for Primary Industries.

Ashbolt NJ, Schoen ME, Soller JA and Roser DJ. (2010). Predicting pathogen risks to aid beach management: The real value of quantitative microbiological risk assessment (QMRA). *Water Research*, 44, 4692-4703.

Atherholt TB, LeChevallier, MW, Norton, WD, Rosen, JS. (1998). Effect of rainfall on Giardia and crypto. *Journal AWWA*, 90, 66-80.

Aw T, Sivaganesan, M, Briggs, S, Dreelin, E, Aslan, A, Dorevitch, S, Shrestha, A, Isaacs, N, Kinzelman, J, Kleinheinz, G, Noble, R, Rediske, R, Scull, B, Rosenberg, S, Weberman, B, Sivy, T, Southwell, B, Siefring, S, Oshima, K, Haugland, R. (2019). Evaluation of multiple laboratory performance and variability in analysis of

recreational freshwaters by a rapid *Escherichia coli* qPCR method (Draft Method C). *Water Research*, 156, 465-474.
Baker M, Kvalsvig, A, Zhang, J, Lake, R, Sears, A, Wilson, N. (2012). Declining Guillain-Barré syndrome after campylobacteriosis control, New Zealand, 1988-2010. *Emerging Infectious Diseases*, 18, 226-233.
Baker MG, Thornley, CN, Lopez, LD, Garrett, NK, Nicol, CM. (2007). A Recurring salmonellosis Epidemic in New Zealand Linked to Contact with Sheep. *Epidemiology and Infection*, 135, 76-83.
Ballantine D, Davies-Colley R. (2013). Nitrogen, phosphorus and *E. coli* loads in the Sherry River, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *47*(4), 529–547
Ballesté E, Blanch, A, Mendez, J, Sala-Comorera, L, Maunula, L, Monteiro, S, Farnleitner, A, Tiehm, A, Jofre, J, García-Aljaro, C. (2021). Bacteriophages are good estimators of human viruses present in water [Original Research]. *Frontiers in Microbiology*, 12, 619495.
Baque R, Gilliam, A, Robles, L, Jakubowski, W, Slifko, T. (2011). A real-time RT-PCR method to detect viable *Giardia lamblia* cysts in environmental waters. *Water Research*, 45, 3175-3184.
Bates D, Mächler, M, Bolker, B, Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1 - 48.
Berumen A, Lennon, R, Breen-Lyles, M, Griffith, J, Patel, R, Boxrud, D, Decuir, M, Farrugia, G, Smith, K, Grover, M. (2021). Characteristics and risk factors of post-infection irritable bowel syndrome after *Campylobacter* enteritis. *Clinical Gastroenterology and Hepatology*, 19, 1855-1863.e1851.
Best E, Powell, E, Swift, C, Grant, K, Frost, J. (2003). Applicability of a rapid duplex real-time PCR assay for speciation of *Campylobacter jejuni* and *Campylobacter coli* directly from culture plates. *FEMS Microbiology Letters*, 229, 237-241.
Bichai F and Smeets PWMH. (2013). Using QMRA-based regulation as a water quality management tool in the water security challenge: Experience from the Netherlands and Australia. *Water Research*, 47, 7315-7326.
Black R, Levine, M, Clements, M, Hughes, T, Blaser, M. (1988). Experimental *Campylobacter jejuni* infection in humans. *Journal of Infectious Diseases*, 157, 472-479.
Blaustein R, Pachepsky, Y, Hill, R, Shelton, D, Whelan, G. (2013). *Escherichia coli* survival in waters: temperature dependence. *Water Research*, 47, 569-578.
Boehm A, Soller, J. (2020). Refined ambient water quality thresholds for human-associated fecal indicator HF183 for recreational waters with and without co-occurring gull fecal contamination. *Microbial Risk Analysis*, 16, 100139.
Boehm AB, Soller, JA, Shanks, OC. (2015). Human-Associated Fecal Quantitative Polymerase Chain Reaction Measurements and Simulated Risk of Gastrointestinal Illness in Recreational Waters Contaminated with Raw Sewage. *Environmental Science & Technology Letters*, 2, 270-275.
Bolton-Ritchie L, Greenfield, S, Madarasz-Smith, A, Milne, J, Stevenson, M, Walker, J. (2013). Recreational water quality: A practitioner's discussion on the limitations of the 2003 national guidelines. *Unpublished paper prepared on behalf of the regional sector's Coastal Special Interest Group (C-SIG) and the Surface Water Integrated Management (SWIM) Group*.
Boucquey D, Sindic, C, Lamy, M, Delmée, M, Tomasi, J, Laterre, E. (1991). Clinical and serological studies in a series of 45 patients with Guillain-Barré syndrome. *Journal of the Neurological Sciences*, 104, 56-63.
Burch T. (2020). Outbreak-based *Giardia* dose–response model using bayesian hierarchical markov chain monte carlo analysis. *Risk Analysis*, 40, 705-722.
Buswell CM, Herlihy YM, Lawrence LM, et al. (1998). Extended survival and persistence of *Campylobacter* spp. in water and aquatic biofilms and their detection by

immunofluorescent-antibody and -rRNA staining. *Applied and Environmental Microbiology*, 64**,** 733-741.
Byappanahalli M, Nevers, M, Korajkic, A, Staley, Z, Harwood, V. (2012). Enterococci in the environment. *Microbiology and Molecular Biology Reviews*, 76, 685-706.
Card T, Enck, P, Barbara, G, Boeckxstaens, G, Santos, J, Azpiroz, F, Mearin, F, Aziz, Q, Marshall, J, Spiller, R. (2018). Post-infectious IBS: defining its clinical features and prognosis using an internet-based survey. *United European Gastroenterology Journal*, 6, 1245-1253.
Carter B, Chalmers, R, Davies, A. (2020). Health sequelae of human cryptosporidiosis in industrialised countries: a systematic review. *Parasites and Vectors*, 13, 443.
Carter PE, McTavish, SM, Brooks, HJL, Campbell, D, Collins-Emerson, JM, Midwinter, AC, French, NP. (2009). Novel clonal complexes with an unknown animal reservoir dominate *Campylobacter jejuni* isolates from river water in New Zealand. *Applied and Environmental Microbiology*, 75, 6038-6046.
Cashdollar J, Brinkman, N, Griffin, S, McMinn, B, Rhodes, E, Varughese, E, Grimm, A, Parshionikar, S, Wymer, L, Fout, G. (2013). Development and evaluation of EPA method 1615 for detection of enterovirus and norovirus in water. *Applied and Environmental Microbiology*, 79, 215-223.
Caudie C, Pinon, A, Taravel, D, Sivadon-Tardy, V, Orlikowski, D, Rozenberg, F, Sharshar, T, Raphaël, J, Gaillard, J. (2011). Preceding infections and anti-ganglioside antibody profiles assessed by a dot immunoassay in 306 French Guillain–Barré syndrome patients. *Journal of Neurology*, 258, 1958-1964.
Chen P, Li, C, Hsieh, T, Chang, C, Lee, H, Lee, N, Wu, C, Lee, C, Shih, H, Ko, W. (2012). Epidemiology, disease spectrum and economic burden of non-typhoidal *Salmonella* infections in Taiwan, 2006–2008. *Epidemiology and Infection*, 140, 2256-2263.
Chern E, Siefring, S, Paar, J, Doolittle, M, Haugland, R. (2011). Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. *Letters in Applied Microbiology*, 52, 298-306.
Chibwe M, Odume, ON, Nnadozie, CF. (2023). Assessment of risk of exposure to *Campylobacter* species and their antibiotic-resistant genes from selected rivers in the Eastern Cape, South Africa. *Environmental Pollution*, 338, 122625.
Clancy JL, Bukhari, Z, McCuin, RM, Matheson, Z, Fricker, CR. (1999). USEPA method 1622. *Journal of the American Water Works Association*, 91, 60-68.
Clancy JL, Gollnitz, WD, Tabib, Z. (1994). Commercial labs: how accurate are they? *Journal of the American Water Works Association*, 86, 89-97.
Connell K, Rodgers, CC, Shank-Givens, HL, Scheller, J, Pope, ML, Miller, K. (2000). Building a better protozoa data set. *Journal of the American Water Works Association*, 92, 30-43.
Cookson A, Burgess, S, Midwinter, A, Marshall, J, Moinet, M, Rogers, L, Fayaz, A, Biggs, P, Brightwell, G. (2024). New *Campylobacter* lineages in New Zealand freshwater: pathogenesis and public health implications. *Environmental Microbiology*, 26, e70016.
Cookson AL, Taylor SCS, Bennett J, Thomson-Carter F, Attwood GT. (2006) Serotypes and analysis of distribution of Shiga toxin-producing *Escherichia coli* from cattle and sheep in the lower North Island, New Zealand. *New Zealand Veterinary* Journal 54(2), 78-84.
Costa D, Iraola, G. (2019). Pathogenomics of Emerging Campylobacter Species. *Clin Microbiol Rev*, 32.
Crank K, Petersen, S, Bibby, K. (2019). Quantitative microbial risk assessment of swimming in sewage impacted waters using CrAssphage and pepper mild mottle virus in a customizable model. *Environmental Science & Technology Letters*, 6, 571-577.
Cressey P. (2021). *Screening quantitative risk assessment: Kaikohe wastewater treatment plant*. CrCfFND Council.
Cressey P, Horn, B. (2016). *New Zealand exposure factors handbook*. Institute of Environmental Science and Research, Christchurch Science Centre, NZ.

Cressey P, Lake, R, Thornley, C. (2014). *Risk ranking: updated estimates of the burden of foodborne disease for New Zealand in 2013*. Ministry for Primary Industries. mpi.govt.nz/dmsdocument/14146/direct/

Cribb D, Varrone, L, Wallace, R, McLure, A, Smith, J, Stafford, R, Bulach, D, Selvey, L, Firestone, S, French, N, Valcanis, M, Fearnley, E, Sloan-Gardner, T, Graham, T, Glass, K, Kirk, M. (2022). Risk factors for campylobacteriosis in Australia: outcomes of a 2018-2019 case-control study. *BMC Infectious Diseases*, 22, 586.

D'Anchino M, Orlando, D, De Feudis, L. (2002). *Giardia lamblia* infections become clinically evident by eliciting symptoms of irritable bowel syndrome. *Journal of Infection*, 45, 169-172.

DeFlorio-Barker S, Arnold, B, Sams, E, Dufour, A, Colford, J, Jr., Weisberg, S, Schiff, K, Wade, T. (2018). Child environmental exposures to water and sand at the beach: findings from studies of over 68,000 subjects at 12 beaches. *Journal of Exposure Science and Environmental Epidemiology*, 28, 93-100.

Delignette-Muller M, Dutang, C. (2015). fitdistrplus: an R package for fitting distributions. *Journal of Statistical Software*, 64, 1 - 34.

Derzelle S, Grine, A, Madic, J, de Garam, C, Vingadassalon, N, Dilasser, F, Jamet, E, Auvray, F. (2011). A quantitative PCR assay for the detection and quantification of Shiga toxin-producing *Escherichia coli* (STEC) in minced beef and dairy products. *International Journal of Food Microbiology*, 151, 44-51.

Devane M, Robson, B, Nourozi, F, Scholes, P, Gilpin, B. (2007). A PCR marker for detection in surface waters of faecal pollution derived from ducks. *Water Research*, 41, 3553-3560.

Devane M, Robson, B, Nourozi, F, Wood, D, Gilpin, B. (2013). Distinguishing human and possum faeces using PCR markers. *Journal of Water and Health*, 11, 397-409.

Devane ML, Nicol, C, Ball, A, Klena, JD, Scholes, P, Hudson, JA, Baker, MG, Gilpin, BJ, Garrett, N, Savill, MG. (2005). The occurrence of *Campylobacter* subtypes in environmental reservoirs and potential transmission routes. *Journal of Applied Microbiology*, 98, 980-990.

DiGiorgio CL, Gonzalez, DA, Huitt, CC. (2002). *Cryptosporidium* and *Giardia* recoveries in natural waters by using Environmental Protection Agency method 1623. *Applied and Environmental Microbiology*, 68, 5952-5955.

Donaldson K, Griffin, D, Paul, J. (2002). Detection, quantitation and identification of enteroviruses from surface waters and sponge tissue from the Florida Keys using real-time RT-PCR. *Water Research*, 36, 2505-2514.

Dufour A, Behymer, T, Cantú, R, Magnuson, M, Wymer, L. (2017). Ingestion of swimming pool water by recreational swimmers. *Journal of Water and Health*, 15, 429-437.

DuPont A. (2008). Postinfectious irritable bowel syndrome. *Clinical Infectious Diseases*, 46, 594-599.

Eisenberg J, Seto, E, Olivieri, A, Spear, R. (1996). Quantifying water pathogen risk in an epidemiological framework. *Risk Analysis*, 16, 549-563.

El-Saadony MT, Saad, AM, Yang, T, Salem, HM, Korma, SA, Ahmed, AE, Mosa, WFA, Abd El-Mageed, TA, Selim, S, Al Jaouni, SK, Zaghloul, RA, Abd El-Hack, ME, El-Tarabily, KA, Ibrahim, SA. (2023). Avian campylobacteriosis, prevalence, sources, hazards, antibiotic resistance, poultry meat contamination, and control measures: a comprehensive review. *Poultry Science*, 102, 102786.

Fasiolo M, Wood, SN, Zaffran, M, Nedellec, R, Goude, Y. (2021). qgam: bayesian nonparametric quantile regression modeling in R. *Journal of Statistical Software*, 100, 1 - 31.

Ferguson C, Kaucner, C, Krogh, M, Deere, D, Warnecke, M. (2004). Comparison of methods for the concentration of *Cryptosporidium* oocysts and *Giardia* cysts from raw waters. *Canadian Journal of Microbiology*, 50, 675-682.

García-Aljaro C, Blanch, A, Campos, C, Jofre, J, Lucena, F. (2019). Pathogens, faecal indicators and human-specific microbial source-tracking markers in sewage. *Journal of Applied Microbiology*, 126, 701-717.

Garcia-R J, Pita, A, Velathanthiri, N, French, N, Hayman, D. (2020). Species and genotypes causing human cryptosporidiosis in New Zealand. *Parasitology Research*, 119, 2317-2326.
Garcia–R J, French, N, Pita, A, Velathanthiri, N, Shrestha, R, Hayman, D. (2017). Local and global genetic diversity of protozoan parasites: Spatial distribution of *Cryptosporidium* and *Giardia* genotypes. *PLOS Neglected Tropical Diseases*, 11, e0005736.
Gibney K, O'Toole, J, Sinclair, M, Leder, K. (2014). Disease burden of selected gastrointestinal pathogens in Australia, 2010. *International Journal of Infectious Diseases* 28, 176-185.
Gilpin B, Hewitt, J, Scholes, P, Coxson, S, Hayman, D, Pita, A, Knox, M, Garcia-Ramirez, J. (2018). *Freshwater microbiological sciences review (FMSR): Stage one. Work stream B: Microbial methods*. ESR Report FW18018. Christchurch.
Gilpin B, Walker, T, Paine, S, Sherwood, J, Mackereth, G, Wood, T, Hambling, T, Hewison, C, Brounts, A, Wilson, M, Scholes, P, Robson, B, Lin, S, Cornelius, A, Rivas, L, Hayman, D, French, N, Zhang, J, Wilkinson, D, Midwinter, A, Biggs, P, Jagroop, A, Eyre, R, Baker, M, Jones, N. (2020). A large scale waterborne Campylobacteriosis outbreak, Havelock North, New Zealand. *Journal of Infection*, 81, 390-395.
Gilpin B, Walshe, G, On, S, Smith, D, Marshall, J, French, N. (2013). Application of molecular epidemiology to understanding campylobacteriosis in the Canterbury region of New Zealand. *Epidemiology and Infection*, 141, 1253-1266.
Gilpin BJ, Scholes, P, Robson, B, Savill, MG. (2008). The Transmission of Thermotolerant Campylobacter spp. to People Living or Working on Dairy Farms in New Zealand. *Zoonoses and Public Health*, 55, 352-360.
Girones R, Bofill-Mas, S. (2013). 22 - Virus indicators for food and water. In N Cook (Ed.), *Viruses in food and water* (pp. 483-509). Woodhead Publishing. https://www.sciencedirect.com/science/article/pii/B9780857094308500228
Green H, Dick, L, Gilpin, B, Samadpour, M, Field, K. (2012). Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. *Applied Environmental Microbiology*, 78, 503-510.
Guy RA, Arsenault, J, Kotchi, SO, Gosselin-Théberge, M, Champagne, M-J, Berthiaume, P. (2018). *Campylobacter* in recreational lake water in southern Quebec, Canada: presence, concentration, and association with precipitation and ruminant farm proximity. *Journal of Water and Health*, 16, 516-529.
Gyawali P, Devane, M, Scholes, P, Hewitt, J. (2021). Application of crAssphage, F-RNA phage and pepper mild mottle virus as indicators of human faecal and norovirus contamination in shellfish. *Science of The Total Environment*, 783, 146848.
Haagsma J, Siersema, P, De Wit, N, Havelaar, A. (2010). Disease burden of post-infectious irritable bowel syndrome in The Netherlands. *Epidemiology and Infection*, 138, 1650-1656.
Halliez M, Buret, A. (2013). Extra-intestinal and long term consequences of *Giardia duodenalis* infections. *World Journal of Gastroenterology*, 19, 8974-8985.
Hannu T, Mattila, L, Rautelin, H, Pelkonen, P, Lahdenne, P, Siitonen, A, Leirisalo-Repo, M. (2002). *Campylobacter*-triggered reactive arthritis: a population-based study. *Rheumatology*, 41, 312-318.
Haugland R, Siefring, S, Lavender, J, Varma, M. (2012). Influences of sample interference and interference controls on quantification of enterococci fecal indicator bacteria in surface water samples by the qPCR method. *Water Research*, 46, 5989-6001.
Health Canada. (2012). *Guidelines for Canadian recreational water quality, Third edition*. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario. (Catalogue No H129-15/2012E).
Health Canada. (2014). *MFLP-46 Isolation of thermophilic campylobacter from food*. Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario.
Henry R, Schang, C, Chandrasena, GI, Deletic, A, Edmunds, M, Jovanovic, D, Kolotelo, P, Schmidt, J, Williamson, R, McCarthy, D. (2015). Environmental monitoring of

waterborne *Campylobacter*: evaluation of the Australian standard and a hybrid extraction-free MPN-PCR method. *Frontiers in Microbiology*, 6, 74.
Hernroth B, Conden-Hansson, A, Rehnstam-Holm, A, Girones, R, Allard, A. (2002). Environmental factors influencing human viral pathogens and their potential indicator organisms in the blue mussel, *Mytilus edulis*: the first Scandinavian report. *Applied Environmental Microbiology*, 68, 4523-4533.
Hewitt J, Bell, D, Simmons, G, Rivera-Aban, M, Wolf, S, Greening, G. (2007). Gastroenteritis outbreak caused by waterborne norovirus at a New Zealand ski resort. *Applied and Environmental Microbiology*, 73, 7853-7857.
Hewitt J, Leonard, M, Greening, G, Lewis, G. (2011). Influence of wastewater treatment process and the population size on human virus profiles in wastewater. *Water Research*, 45, 6267-6276.
Highton J, Priest, D. (1996). Reactive arthritis: characteristics in southern New Zealand. *New Zealand Medical Journal*, 109, 93-95.
Hill V, Polaczyk, A, Hahn, D, Narayanan, J, Cromeans, T, Roberts, J, Amburgey, J. (2005). Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants. *Applied and Environmental Microbiology*, 71, 6878-6884.
Hoorfar J, Ahrens, P, Rådström, P. (2000). Automated 5' nuclease PCR assay for identification of *Salmonella enterica*. *Journal of Clinical Microbiology*, 38, 3429-3435.
Hoque M, Hope, V, Scragg, R, Kjellström, T. (2003). Children at risk of giardiasis in Auckland: a case-control analysis. *Epidemiology and Infection*, 131, 655-662.
Horn B, McBride, G, Lake, R, Muirhead, R. (2018). *Freshwater Microbiological Sciences Review (FMSR):Stage one work stream C: Dose response illness and infection risk*. ESR Report FW17068. Christchurch
Horn B, Wright, J, Rivas, L, Cressey, P, Armstrong, B, Paine, S, Winter, D, Wang, J, Vidovic, S, Mills, J, A-L., H, Hazard, J, Ware, L, Soboleva, T, Thomas, K, Gibbs, M. (2024). *Attribution and characterisation of New Zealand domestic shiga toxin-producing Escherichia coli infections*.
Hsu B-M, Huang, C. (2000). Recovery of *Giardia* and *Cryptosporidium* from water by various concentration, elution, and purification techniques. *Journal of Environmental Quality*, 29, 1587-1593.
Hsu B-M, Huang, C, Yeh, H-H. (2002). Evaluation of the ICR protozoan method and Method 1623 for detecting Giardia and Cryptosporidium in actual water samples. *Water Supply*, 2, 89-93.
Hughes R, Cornblath, D. (2005). Guillain-Barré syndrome. *Lancet*, 366, 1653-1666.
Ibekwe A, Watt, P, Shouse, P, Grieve, C. (2004). Fate of *Escherichia coli* O157:H7 in irrigation water on soils and plants as validated by culture method and real-time PCR. *Canadian Journal of Microbiology*, 50, 1007-1014.
Iglói Z, Mughini-Gras, L, Nic Lochlainn, L, Barrasa, A, Sane, J, Mooij, S, Schimmer, B, Roelfsema, J, van Pelt, W, Kortbeek, T. (2018). Long-term sequelae of sporadic cryptosporidiosis: a follow-up study. *European Journal of Clinical Microbiology and Infectious Diseases*, 37, 1377-1384.
Igwaran A, Okoh, AI. (2019). Human campylobacteriosis: A public health concern of global importance. *Heliyon*, 5, e02814.
International Organization for Standardization. (2012). *Microbiology of food and animal feed – real-time polymerase chain reaction (PCR)-based method for the detection of food-borne pathogens - horizontal method for the detection of shiga toxin-producing Escherichia coli (STEC) and the determination of O157, O111, O26, O103 and O143 serogroups*. International Organization for Standardization, Geneva.
International Organization for Standardization. (2017). *ISO 10272-1:2017 microbiology of the food chain - horizontal method for detection and enumeration of Campylobacter spp. — Part 1: detection method*. International Organization for Standardization, Geneva.

Ipsos Public Affairs. (2020). *Department of Conservation Survey of New Zealanders 2020*. Department of Conservation. https://www.doc.govt.nz/globalassets/documents/about-doc/role/visitor-research/survey-of-new-zealanders-2020.pdf.
Jackson B, Zegarra, J, López-Gatell, H, Sejvar, J, Arzate, F, Waterman, S, NÚÑEz, AS, LÓPez, B, Weiss, J, Cruz, R, Murrieta, D, Luna-Gierke, R, Heiman, K, Vieira, A, Fitzgerald, C, Kwan, P, Zárate-Bermúdez, M, Talkington, D, Hill, V, Mahon, B. (2014). Binational outbreak of Guillain–Barré syndrome associated with *Campylobacter jejuni* infection, Mexico and USA, 2011. *Epidemiology and Infection*, 142, 1089-1099.
Jacobs B, Rothbarth, P, van der Meché, F, Herbrink, P, Schmitz, P, de Klerk, M, van Doorn, P. (1998). The spectrum of antecedent infections in Guillain-Barré syndrome: a case-control study. *Neurology*, 51, 1110-1115.
Jahne M, Schoen, M, Garland, J, Nappier, S, Soller, J. (2024). Microbial treatment targets for potable and nonpotable water reuse – a comprehensive update and harmonization. *Environmental Science and Technology Letters*, 11, 1175-1181.
Jolley K, Bray, J, Maiden, M. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Research*, 3, 124.
Kageyama T, Kojima, S, Shinohara, M, Uchida, K, Fukushi, S, Hoshino, F, Takeda, N, Katayama, K. (2003). Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *Journal of Clinical Microbiology*, 41, 1548-1557.
Kageyama T, Shinohara, M, Uchida, K, Fukushi, S, Hoshino Fuminori, B, Kojima, S, Takai, R, Oka, T, Takeda, N, Katayama, K. (2004). Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to Norovirus in Japan. *Journal of Clinical Microbiology*, 42, 2988-2995.
Kay D, Fleisher JM, Salmon RL, et al. (1994). Predicting likelihood of gastroenteritis from sea bathing: resutls from randomised exposure. *The Lancet*, 344, 905-909.
Keithlin J, Sargeant, J, Thomas, M, Fazil, A. (2014). Systematic review and meta-analysis of the proportion of *Campylobacter* cases that develop chronic sequelae. *BMC Public Health*, 14, 1203.
Kemmeren J, Mangen, M-J, van Duynhoven, Y, Havelaar, A. (2006). *Priority setting of foodborne pathogens: disease burden and costs of selected enteric pathogens*. RIVM Report 330080001/2006. https://www.rivm.nl/bibliotheek/rapporten/330080001.pdf.
Kent G, Greenspan, J, Herndon, J, Mofenson, L, Harris, J, Eng, T, Waskin, H. (1988). Epidemic giardiasis caused by a contaminated public water supply. *American Journal of Public Health*, 78, 139-143.
Kingsbury J, Vather, N, Thomas, K. (2024). *Risk profile update: Salmonella (nontyphoidal) in and on eggs*. Ministry for Primary Industries https://www.mpi.govt.nz/dmsdocument/61102/direct/.
Kinzelman J, Bushon, R, Dorevitch, S, Noble, R. (2011). *Comparative evaluation of molecular and culture methods for fecal indicator bacteria for use in inland recreational waters*. IWA Publishing. https://doi.org/10.2166/9781843395454.
Kirk M, Pires, S, Black, R, Caipo, M, Crump, J, Devleesschauwer, B, Döpfer, D, Fazil, A, Fischer-Walker, C, Hald, T, Hall, A, Keddy, K, Lake, R, Lanata, C, Torgerson, P, Havelaar, A, Angulo, F. (2015). World Health Organization etimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Medicine*, 12, e1001921.
Kirner J, Littler, J, Angelo, L. (1978). A waterborne outbreak of giardiasis in Camas, Wash. *Journal American Water Works Association*, 70, 35-40.
Kobayashi M, Zhang, Q, Segawa, T, Maeda, M, Hirano, R, Okabe, S, Ishii, S. (2022). Temporal dynamics of *Campylobacter* and *Arcobacter* in a freshwater lake that receives fecal inputs from migratory geese. *Water Research*, 217, 118397.
Koenker R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.

Koga M, Ang, C, Yuki, N, Jacobs, B, Herbrink, P, van der Meché, F, Hirata, K, van Doorn, P. (2001). Comparative study of preceding *Campylobacter jejuni* infection in Guillain-Barré syndrome in Japan and The Netherlands. *Journal of Neurology, Neurosurgery and Psychiatry*, 70, 693-695.

Korajkic A, McMinn, B, Harwood, V. (2018). Relationships between microbial indicators and pathogens in recreational water settings. *International Journal of Environmental Research and Public Health*, 15, 2842.

Korajkic A, Wanjugi, P, Harwood, V. (2013). Indigenous microbiota and habitat influence *Escherichia coli* survival more than sunlight in simulated aquatic environments. *Applied and Environmental Microbiology*, 79, 5329-5337.

Kukkula M, Arstila, P, Klossner, M, Maunula, L, Bonsdorff, C, Jaatinen, P. (1997). Waterborne outbreak of viral gastroenteritis. *Scandinavian Journal of Infectious Diseases*, 29, 415-418.

Lake R, Campbell, D, Hathaway, S, Ashmore, E, Cressey, P, Horn, B, Pirikahu, S, Sherwood, J, Baker, M, Shoemack, P, Benschop, J, Marshall, J, Midwinter, A, Wilkinson, D, French, N. (2021). Source attributed case-control study of campylobacteriosis in New Zealand. *International Journal of Infectious Diseases*, 103, 268-277.

Lake R, Horn, B, McBride, G, Muirhead, R. (2018). *Freshwater microbiological sciences review (FMSR): Stage one work stream C: Dose response recommendation for dose response*. ESR Report FW17069. Christchurch.

Leonard M. (2025) Review for scoping water quality. *Guidelines for recreational water and shellfish gathering.* ESR Report FW24043. Christchurch.

Leonard M, Eaton C. (2021) *Recreational Water Quality Guidelines update*. ESR Report FW21029. Christchurch.

Leonard M, Gilpin, B, Horn, B, Coxson, S, Armstrong, B, Hewitt, J, Devane, M, Gyawali, P, Scholes, P, Thom, K, Cornelius, A, Orton, A, Hayman, D, Pita, A. (2020). *Quantitative microbial risk assessment pilot*. ESR Report FW200007. Christchurch.

Litleskare S, Rortveit, G, Eide, G, Hanevik, K, Langeland, N, Wensaas, K. (2018). Prevalence of irritable bowel syndrome and chronic fatigue 10 years after *Giardia* infection. *Clinical Gastroenterology and Hepatology* 16, 1064-1072.e1064.

Lynch JP, 3rd, Kajon, AE. (2016). Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Semin Respir Crit Care Med*, 37, 586-602.

Mafumo N, Bezuidt, OKI, le Roux, W, Makhalanyane, TP. (2023). CrAssphage May Be Viable Markers of Contamination in Pristine and Contaminated River Water. *mSystems*, 8, e0128222.

Maiden M, Bygraves, J, Feil, E, Morelli, G, Russell, J, Urwin, R, Zhang, Q, Zhou, J, Zurth, K, Caugant, D, Feavers, I, Achtman, M, Spratt, B. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Science USA*, 95, 3140-3145.

Majowicz S, Panagiotoglou, D, Taylor, M, Gohari, M, Kaplan, G, Chaurasia, A, Leatherdale, S, Cook, R, Patrick, D, Ethelberg, S. (2020). Determining the long-term health burden and risk of sequelae for 14 foodborne infections in British Columbia, Canada: protocol for a retrospective population-based cohort study. *BMJ Open*, 10, e036560.

Malorny B, Paccassoni, E, Fach, P, Bunge, C, Martin, A, Helmuth, R. (2004). Diagnostic real-time PCR for detection of *Salmonella* in food. *Applied and Environmental Microbiology*, 70, 7046-7052.

Mangen M-J, Bouwknegt, M, Friesema, I, Haagsma, J, Kortbeek, L, Tariq, L, Wilson, M, van Pelt, W, Havelaar, A. (2015). Cost-of-illness and disease burden of food-related pathogens in the Netherlands, 2011. *International Journal of Food Microbiology*, 196, 84-93.

Mary C, Chapey, E, Dutoit, E, Guyot, K, Hasseine, L, Jeddi, F, Menotti, J, Paraud, C, Pomares, C, Rabodonirina, M, Rieux, A, Derouin, F. (2013). Multicentric evaluation of

a new real-time PCR assay for quantification of *Cryptosporidium* spp. and identification of *Cryptosporidium parvum* and *Cryptosporidium hominis*. *Journal of Clinical Microbiology*, 51, 2556-2563.
Matsuki T, Watanabe, K, Fujimoto, J, Kado, Y, Takada, T, Matsumoto, K, Tanaka, R. (2004). Quantitative PCR with 16S rRNA-gene-targeted species-specific primers for analysis of human intestinal bifidobacteria. *Applied Environmental Microbiology*, 70, 167-173.
McBride G. (2005). MPNs and microbiology. In *Using Statistical Methods for Water Quality Management: Issues, problems and solutions*. John Wiley & Sons.
McBride G, Thorn, C, Salmond, C. (1996). *Feasibility of bathing-health effects study for New Zealand freshwaters*. Report prepared by NIWA for Ministry for the Environment.
McBride G, Till, D, Ryan, T, Ball, A, Lewis, G, Palmer, S, Weinstein, P. (2002). *Freshwater microbiology research programme report: pathogen occurrence and human health risk assessment analysis*. Ministry for the Environment, Wellington.
McBride G, Soller J. (2017). *Technical background for 2017 MfE 'clean water' swimmability proposals for rivers.* National Institute of Water and Atmospheric Research. FWWQ1722.
McCarthy N, Giesecke, J. (2001). Incidence of Guillain-Barré syndrome following infection with *Campylobacter jejuni*. *American Journal of Epidemiology*, 153, 610-614.
McCuin RM, Clancy, JL. (2003). Modifications to United States Environmental Protection Agency methods 1622 and 1623 for detection of *Cryptosporidium* oocysts and *Giardia* cysts in water. *Applied and Environmental Microbiology*, 69, 267-274.
Mearin F, Pérez-Oliveras, M, Perelló, A, Vinyet, J, Ibañez, A, Coderch, J, Perona, M. (2005). Dyspepsia and irritable bowel syndrome after a *Salmonella* gastroenteritis outbreak: one-year follow-up cohort study. *Gastroenterology*, 129, 98-104.
Medema G, Bahar, M, Schets, F. (1997). Survival of *Cryptosporidium parvum*, *Escherichia coli*, faecal enterococci and *Clostridium perfringens* in river water: influence of temperature and autochthonous microorganisms. *Water Science and Technology*, 35, 249-252.
Medema G, Teunis, P, Havelaar, A, Haas, C. (1996). Assessment of the dose-response relationship of *Campylobacter jejuni*. *International Journal of Food Microbiology*, 30, 101-111.
Messner M, Berger, P. (2016). *Cryptosporidium* infection risk: results of new dose-response modeling. *Risk Analysis*, 36, 1969-1982.
Messner MJ, Berger, P. (2016). Infection Risk: Results of New Dose-Response Modeling. *Risk Analysis*, 36, 1969-1982.
Milne J, Madarasz-Smith, A, Davies, T. (2017). *Recreational water quality monitoring and reporting in New Zealand: A discussion paper for Regional and Unitary Councils.*
Milne J, Moriarty, E, Davie, T, Kitson, J, Gilles, T. (2018). *Freshwater microbiological sciences review (FMSR): Stage one. Work stream A: site selection. Recommended sampling sites.* ESR Report CSC18004. Christchurch
MIMM. (2008). Method 13.1 thermotolerant Campylobacters: MPN method. In *Meat industry microbiological methods 5th ed. Online*.
Ministry for the Environment. (2003). *Microbiological water quality guidelines for marine and freshwater recreational areas*
Mishu B, Blaser, M. (1993). Role of infection due to *Campylobacter jejuni* in the initiation of Guillain-Barre syndrome. *Clinical Infectious Diseases*, 17, 104-108.
Mishu B, Ilyas, A, Koski, C, Vriesendorp, F, Cook, S, Mithen, F, Blaser, M. (1993). Serologic evidence of previous *Campylobacter jejuni* infection in patients with the Guillain-Barré syndrome. *Annals of Internal Medicine*, 118, 947-953.
Moran K. (2009). *Re-thinking drowning risk, the role of water safety knowledge, attitudes, and behaviours in the aquatic recreation of New Zealand youth*. Riga, Latvia: VDM Verlag.
Moriarty E, Milne, J, Gilpin, B, McBride, G, Horn, B, Muirhead, R, Lake, R, Hayman, D, Coxson, S, Pita, A. (2018). *Freshwater microbiological sciences review (stage one).*

Work stream D: Survey design and QMRA project plan. ESR Report CSC18006. Christchurch.
Moss-Morris R, Spence, M. (2006). To "lump" or to "split" the functional somatic syndromes: can infectious and emotional risk factors differentiate between the onset of chronic fatigue syndrome and irritable bowel syndrome? *Psychosomatic medicine*, 68, 463-469.
Müllner P, Collins-Emerson, JM, Midwinter, AC, Carter, P, Spencer, SE, van der Logt, P, Hathaway, S, French, NP. (2010). Molecular epidemiology of Campylobacter jejuni in a geographically isolated country with a uniquely structured poultry industry. *Appl Environ Microbiol*, 76, 2145-2154.
New Zealand Government. (2020). *National policy statement for freshwater management 2020*. Water services (drinking water standards for New Zealand) regulations 2022, (2022).
Nygård K, Schimmer, B, Søbstad, Ø, Walde, A, Tveit, I, Langeland, N, Hausken, T, Aavitsland, P. (2006). A large community outbreak of waterborne giardiasis-delayed detection in a non-endemic urban area. *BMC Public Health*, 6, 141.
Pachepsky Y, Blaustein, R, Whelan, G, Shelton, D. (2014). Comparing temperature effects on *Escherichia coli*, *Salmonella*, and *Enterococcus* survival in surface waters. *Letters in Applied Microbiology*, 59, 278-283.
Paton A, Paton, J. (1998). Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for *stx*$_1$, *stx*$_2$, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfb*$_{O111}$, and *rfb*$_{O157}$. *Journal of Clinical Microbiology*, 36, 598-602.
Pattis I, Horn B, Kingsbury J, Cressey P, L, L. (2025). *Annual report concerning foodborne disease in New Zealand , 2024.* ESR client report for Ministry for Primary Industries.
Petruzziello-Pellegrini T, Marsden, P. (2012). Shiga toxin-associated hemolytic uremic syndrome: advances in pathogenesis and therapeutics. *Current Opinion in Nephrology and Hypertension*, 21, 433-440.
Phiri BJ, Pita, AB, Hayman, DTS, Biggs, PJ, Davis, MT, Fayaz, A, Canning, AD, French, NP, Death, RG. (2020). Does land use affect pathogen presence in New Zealand drinking water supplies? *Water Research*, 185, 116229.
Pires S, Jakobsen, L, Ellis-Iversen, J, Pessoa, J, Ethelberg, S. (2019). Burden of disease estimates of seven pathogens commonly transmitted through foods in Denmark, 2017. *Foodborne Pathogens and Disease*, 17, 322-339.
Pogreba-Brown K, Austhof, E, Tang, X, Trejo, M, Owusu-Dommey, A, Boyd, K, Armstrong, A, Schaefer, K, Bazaco, M, Batz, M, Riddle, M, Porter, C. (2021). Enteric pathogens and reactive arthritis: systematic review and meta-analyses of pathogen-associated reactive arthritis. *Foodborne Pathogens and Disease*, 18, 627-639.
Pond K. (2013). *Water recreation and disease: plausibility of associated infections: acute effects, sequelae, and mortality*. London: IWA Publishing.
Porcari S, Ingrosso, M, Maida, M, Eusebi, L, Black, C, Gasbarrini, A, Cammarota, G, Ford, A, Ianiro, G. (2024). Prevalence of irritable bowel syndrome and functional dyspepsia after acute gastroenteritis: systematic review and meta-analysis. *Gut*, 73, 1431-1440.
Pouillot R, Delignette-Muller, M. (2010). Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology*, 142, 330-340.
Powell MJD. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Report No. DAMTP 2009/NA06, Centre for Mathematical Sciences, University of Cambridge, UK. http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf.
Public Health England. (2015). *Identification of Salmonella species. ID 24 Issue 3. UK standards for microbiology investigations*.
Puig M, Jofre, J, Lucena, F, Allard, A, Wadell, G, Girones, R. (1994). Detection of adenoviruses and enteroviruses in polluted waters by nested PCR amplification. *Applied Environmental Microbiology*, 60, 2963-2970.

R Core Team. (2024). R: A Language and Environment for Statistical Computing (Version 4.2.0) [Computer software].
Rapp D, Ross, CM, Cave, V, Muirhead, RW. (2014). Prevalence, concentration and genotypes of Campylobacter jejuni in faeces from dairy herds managed in farm systems with or without housing. *Journal of Applied Microbiology*, 116, 1035-1043.
Ravel A, Nesbitt, A, Pintar, K, Macarthur, A, Wang, H, Marshall, B, Pollari, F. (2013). Epidemiological and clinical description of the top three reportable parasitic diseases in a Canadian community. *Epidemiology and Infection*, 141, 431-442.
Rees J, Soudain, S, Gregson, N, Hughes, R. (1995). *Campylobacter jejuni* infection and Guillain-Barré syndrome. *New England Journal of Medicine*, 333, 1374-1379.
Reischer G, Kasper, D, Steinborn, R, Mach, R, Farnleitner, A. (2006). Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. *Applied and Environmental Microbiology*, 72, 5610-5614.
Rendtorff R. (1954). The experimental transmission of human intestinal protozoan parasites. II. *Giardia lamblia* cysts given in capsules. *American Journal of Hygiene*, 59, 209-220.
Rendtorff R, Holt, C. (1954). The experimental transmission of human intestinal protozoan parasites. IV. Attempts to transmit *Endamoeba coli* and *Giardia lamblia* cysts by water. *American Journal of Hygiene*, 60, 327-338.
Rivas L, Horn, B, Armstrong, B, Wright, J, Strydom, H, Wang, J, Paine, S, Thom, K, Orton, A, Robson, B, Lin, S, Wong, J, Brunton, C, Smith, D, Cooper, J, Mangalasseril, L, Thornley, C, Gilpin, B. (2024). A case-control study and molecular epidemiology of yersiniosis in Aotearoa New Zealand. *Journal of Clinical Microbiology*, 62, e0075424.
Rose J, Gerba, C. (1991). Use of risk assessment for development of microbial standards. *Water Science and Technology*, 24, 29-34.
Rose J, Haas, C, Regli, S. (1991). Risk assessment and control of waterborne giardiasis. *American Journal of Public Health*, 81, 709-713.
Scallan Walter E, Crim, S, Bruce, B, Griffin, P. (2019). Postinfectious irritable bowel syndrome after *Campylobacter* infection. *American Journal of Gastroenterology*, 114, 1649-1656.
Schets F, Schijven, J, de Roda Husman, A. (2011). Exposure assessment for swimmers in bathing waters and swimming pools. *Water Research*, 45, 2392-2400.
Schijven JF, de Roda Husman, A. (2006). A survey of diving behaviour and accidental water ingestion among Dutch occupational and sport divers to assess the risk of infection with waterborne pathogenic microorganisms. *Environmental Health Perspectives*, 114, 712-717.
Schijven JF, Teunis PFM, Rutjes SA, Bouwknegt M and de Roda Husman A. (2011) QMRAspot: A tool for Quantitative Microbiological Risk Assessment from surface water to potable water. *Water Reseach*, 45, 5564-5576.
Schoen ME, Boehm, AB, Soller, J, Shanks, OC. (2020). Contamination scenario matters when using viral and bacterial human-associated genetic markers as indicators of a health risk in untreated sewage-impacted recreational waters. *Environmental Science & Technology*, 54, 13101-13109.
Schönberg-Norio D, Mattila, L, Lauhio, A, Katila, M, Kaukoranta, S, Koskela, M, Pajarre, S, Uksila, J, Eerola, E, Sarna, S, Rautelin, H. (2010). Patient-reported complications associated with *Campylobacter jejuni* infection. *Epidemiology and Infection*, 138, 1004-1011.
Schwille-Kiuntke J, Enck, P, Zendler, C, Krieg, M, Polster, A, Klosterhalfen, S, Autenrieth, I, Zipfel, S, Frick, J. (2011). Postinfectious irritable bowel syndrome: follow-up of a patient cohort of confirmed cases of bacterial infection with *Salmonella* or *Campylobacter*. *Neurogastroenterology and Motility*, 23, e479-488.
Scott TM, Rose, JB, Jenkins, TM, Farrah, SR, Lukasik, J. (2002). Microbial source tracking: current methodology and future directions. *Appl Environ Microbiol*, 68, 5796-5803.

Seemann T, Goncalves, D, Bulach, D, Schultz, M, Kwong, J, Howden, B. (n. d.). Nullarbor github. https://github.com/tseemann/nullarbor.
Shanks O, White, K, Kelty, C, Sivaganesan, M, Blannon, J, Meckes, M, Varma, M, Haugland, R. (2010). Performance of PCR-based assays targeting *Bacteroidales* genetic markers of human fecal pollution in sewage and fecal samples. *Environmental Science and Technology*, 44, 6281-6288.
Sharma A, Lal, V, Modi, M, Vaishnavi, C, Prabhakar, S. (2011). *Campylobacter jejuni* infection in Guillain-Barré syndrome: a prospective case control study in a tertiary care hospital. *Neurology India*, 59, 717-721.
Shrestha R, Midwinter, A, Marshall, J, Collins-Emerson, J, Pleydell, E, French, N. (2019). *Campylobacter jejuni* strains associated with wild birds and those causing human disease in six high-use recreational waterways in New Zealand. *Applied and Environmental Microbiology*, 85, e01228-01219.
Siegler R, Christofferson, R, Milligan, M, Pavia, A. (1994). A 20-year population-based study of postdiarrheal hemolytic uremic syndrome in Utah. *Pediatrics*, 94, 35-40.
Sinclair R, Jones, E, Gerba, C. (2009). Viruses in recreational water-borne disease outbreaks: a review. *Journal of Applied Microbiology*, 107, 1769-1780.
Skiendzielewski K, Burch, T, Stokdyk, J, McGinnis, S, McLoughlin, S, Firnstahl, A, Spencer, S, Borchardt, M, Murphy, H. (2024). Two risk assessments: evaluating the use of indicator HF183 *Bacteroides* versus pathogen measurements for modelling recreational illness risks in an urban watershed. *Water Research*, 259, 121852.
Soller AJ, Scheon ME, Bartrand T,Ravenscroft JE and Ashbolt NJ. (2010) Estimated human health risks from exposure to recreational water impacted by human and non-human sources of faecal contamination. Water Research, 44, 4674-4691.
Soller AJ, Schoen ME, Varghese A, Ichida AM, Boehm AB, Eftim S, Ashbolt NJ and Ravenscroft JE. (2014). Human health risk implications of mulitple sources of feacal indicator bacteria in a recreational waterbody. *Water Research*, 66, 254-264.
Soller AJ, Eftim S, Wade TJ et al. (2016). Use of quantitative microbiological risk assessment to improve interpretation of a recreational water epidemiological study. Microbial Risk Analysis,1, 2-11.
Stidson RT, Gray CA and McPhail CD. (2012). Development and use of modelling techniques for real-time bathing water quality predictions. *Water and Environment Journal*, 26, 7-18.
Stiff RE, Davies, AP, Mason, BW, Hutchings, HA, Chalmers, RM. (2017). Long-term health effects after resolution of acute *Cryptosporidium parvum* infection: a 1-year follow-up of outbreak-associated cases. *Journal of Medical Microbiology*, 66, 1607-1611.
Svendsen A, Bytzer, P, Engsbro, A. (2019). Systematic review with meta-analyses: does the pathogen matter in post-infectious irritable bowel syndrome? *Scandinavian Journal of Gastroenterology*, 54, 546-562.
Tam C, O'Brien, S, Petersen, I, Islam, A, Hayward, A, Rodrigues, L. (2007). Guillain-Barré syndrome and preceding infection with campylobacter, influenza and Epstein-Barr virus in the general practice research database. *PLoS One*, 2, e344.
Tam C, Rodrigues, L, Petersen, I, Islam, A, Hayward, A, O'Brien, S. (2006). Incidence of Guillain-Barré syndrome among patients with *Campylobacter* infection: a general practice research database study. *Journal of Infectious Diseases*, 194, 95-97.
Tam C, Rodrigues, L, Viviani, L, Dodds, J, Evans, M, Hunter, P, Gray, J, Letley, L, Rait, G, Tompkins, D, O'Brien, S. (2012). Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*, 61, 69-77.
Tang W, Mohandas, J, McDonald, S, Hawley, C, Badve, S, Boudville, N, Brown, F, Clayton, P, Wiggins, K, Bannister, K, Campbell, S, Johnson, D. (2012). End-stage kidney disease due to haemolytic uraemic syndrome – outcomes in 241 consecutive ANZDATA registry cases. *BMC Nephrology*, 13, 164.

Tani N, Dohi, Y, Kurumatani, N, Yonemasu, K. (1995). Seasonal distribution of adenoviruses, enteroviruses and reoviruses in urban river water. *Microbiology and Immunology*, 39, 577-580.
Till, D.; McBride, G.; Ball, A.; Taylor, K.; Pyle, E. (2008). Large-scale freshwater microbiological study: Rationale, results and risks. *Journal of Water and Health 6(4)*: 443–460.
Teunis P, Marinović, A, Tribble, D, Porter, C, Swart, A. (2018). Acute illness from *Campylobacter jejuni* may require high doses while infection occurs at low doses. *Epidemics*, 24, 1-20.
United States Environmental Protection Agency. (2005). *Method 1623: Cryptosporidium and Giardia in water by filtration/IMS/FA*. Washington DC: United States Environmental Protection Agency
United States Environmental Protection Agency. (2012). *2012 recreational water quality criteria. Report 820-F-12-058*. Washington DC: United States Environmental Protection Agency. https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf
United States Environmental Protection Agency. (2014). *Site-specific alternative recreational criteria technical support materials for alternative indicators and methods*. Report EPA-820-R-14-011 Washington DC: United States Environmental Protection Agency. https://www.epa.gov/sites/default/files/2015-11/documents/sitespecific-alternative-recreational-indicators-methods.pdf.
United States Environmental Protection Agency. (2024). *Technical support materials: developing alternative recreational criteria for waters contaminated by predominantly non-human fecal sources*. Washington DC: United States Environmental Protection Agency https://www.epa.gov/system/files/documents/2024-07/tsm-nonhuman-sources-revised_073024_508c.pdf.
USEPA. (1996). *Evaluation of the Proposed ICR Protocol for Determining the Field Recovery of Protozoan Cysts From Source and Treated Drinking Water. Final Report.*
Vose D. (2008). *Risk analysis: a quantitative guide*. Hoboken, New Jersey: John Wiley & Sons Ltd.
Wade TJ, Arnold BF, Schiff K, Colford JM Jr., Weisberg SB, Griffith JF, et al. (2022) Health risks to children from exposure to fecally contaminatedrecreational water. *PLoS ONE* 17(4): e0266749.
Walker T, Grainger, R, Quirke, T, Roos, R, Sherwood, J, Mackereth, G, Kiedrzynski, T, Eyre, R, Paine, S, Wood, T, Jagroop, A, Baker, M, Jones, N. (2022). Reactive arthritis incidence in a community cohort following a large waterborne campylobacteriosis outbreak in Havelock North, New Zealand. *BMJ Open*, 12, e060173.
Wang X, Yu, D, Chui, L, Zhou, T, Feng, Y, Cao, Y, Zhi, S. (2024). A Comprehensive Review on Shiga Toxin Subtypes and Their Niche-Related Distribution Characteristics in Shiga-Toxin-Producing E. coli and Other Bacterial Hosts. *Microorganisms*, 12.
Wiedenmann A, Krüger P, Dietz K, Lopez-Pila J, Szewzyk R and Botzenhart K. (2006). Randomized controlled trial assessing infectious disease risks from bathing in fresh recreational waters in relation to the concentration of *Escherichia coli*, intestinal enterococci, *Clostridium perfringens*, and somatic coliphages. *Environmental Health Perspectives*, 114(2), 228‑36.
Wilkinson S, Davy, S, Bunce, M, Stat, M. (2018). Taxonomic identification of environmental DNA with informatic sequence classification trees. *PeerJ Preprints*, 6, e26812v26811.
Wolf S, Hewitt, J, Greening, G. (2010). Viral multiplex quantitative PCR assays for tracking sources of fecal contamination. *Applied and Environmental Microbiology*, 76, 1388-1394.
Wong T, Devane, ML, Hudson, JA, Scholes, P, Grace, SM, Klena, JD. (2004). Validation of a PCR method for *Campylobacter* detection on poultry packs. *British Food Journal*, 106, 642-650.

Wong W, Prestidge, C, Dickens, A, Ronaldson, J. (2023). Diarrhoea-associated haemolytic uraemic syndrome and Shiga toxin-producing *Escherichia coli* infections in New Zealand children: Clinical features and short-term complications from a 23-year cohort study. *Journal of Paediatrics and Child Health*, 59, 493-498.

Wood SN. (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC.

World Health Organization. (1999). *Health-based monitoring of recreational waters: the feasibility of a new approach (the "Annapolis Protocol")*. https://iris.who.int/handle/10665/66477.

World Health Organization. (2003). *Guidelines for safe recreational water environments. Volume 1, Coastal and fresh waters*. https://www.who.int/publications/i/item/9241545801.

World Health Organisation. (2009). Addendum to WHO *Guidelines for safe recreational water environments. Volume 1, Coastal and fresh waters*. https://iris.who.int/handle/10665/70226.

World Health Organization. (2015). *WHO estimates of the global burden of foodborne diseases: foodborne disease burden epidemiology reference group 2007-2015*. https://www.who.int/publications/i/item/9789241565165.

World Health Organization. (2016). *Quantitative microbial risk assessment: application for water safety management*. https://www.who.int/publications/i/item/9789241565370.

World Health Organization. (2021). *Guidelines on recreational water quality. Volume 1: coastal and fresh waters*. https://www.who.int/publications/i/item/9789240031302.

Wyer MD, Kay D, Morgan H et al. (2013) Statisitcal modelling of faecal inidcator organisms at a marine bathing water site: results of an intensive study at Swansea Bay, UK. Report from the Interreg 4a Smart Coasts – Sustainable Communities Project. Aberystwyth University.

Yoneda I, Nishiyama, M, Watanabe, T. (2024). Significant factors for modelling survival of *Escherichia coli* in lake sediments. *Microorganisms*, *12*(6), 1192.

Youn Y, Kim, H, Lim, H, Park, J, Kim, J, Park, H. (2016). Long-term clinical course of post-infectious irritable bowel syndrome after shigellosis: a 10-year follow-up study. *Journal of Neurogastroenterology and Motility*, 22, 490-496.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA and Smith GM. (2009). Mixed effects models and extensions in ecology with R. Statistics for Biology and Health Series, Springer.