

ESTIMATING PARAMETER DISTRIBUTIONS AND DEVELOPING EPIDEMIOLOGICAL METHODS AND TOOLS TO SUPPORT GASTROINTESTINAL OUTBREAK INVESTIGATION

Adedoyin Abiola Awofisayo-Okuyelu

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Abstract

Incubation periods of common gastrointestinal pathogens are cited inconsistently without a solid evidence base for these distributions or factors affecting incubation time. In some high-profile outbreaks, such as the STEC O104 outbreak in Germany or the large 2009 STEC O157 outbreak in South East England, incubation periods were longer than expected complicating the collection of exposure data and the outbreak investigation process. This research aimed to use observational and experimental data to understand the incubation period distribution of some common gastrointestinal pathogens and identify factors that influence these distributions.

I conducted systematic reviews on the incubation period of *Campylobacter*, STEC and *Salmonella* Typhi. In addition to using traditional methods for testing and describing the pattern of heterogeneity, I developed methods for analysing data from observational studies which involved using Kolmogorov-Smirnov test and hierarchical cluster analysis. I also extracted and analysed individual patient data of cases associated with outbreaks of STEC reported in England and Wales. Finally, I developed a within-host mathematical model that

simulated the process of *Salmonella* Typhi infection and conducted a sensitivity analysis to identify influencing parameters.

The systematic reviews and meta-analysis of individual patient data indicated the presence of heterogeneity across outbreaks which could not be explained by chance alone. Some factors associated with the distribution of incubation period were age, food vehicle, mode of transmission, attack rate and vaccine history. The mathematical model estimated an incubation period of 9.6 days for *Salmonella* Typhi, and parameters strongly correlated with incubation period included the net population of bacteria in the colon and bacterial growth during systemic infection.

Outputs of this research will inform the development of policies and public health interventions that contribute towards reducing the burden of disease due to gastrointestinal infections. Further research, developing a stochastic model of the incubation period of *Salmonella* Typhi, is recommended.

**ESTIMATING PARAMETER DISTRIBUTIONS AND
DEVELOPING EPIDEMIOLOGICAL METHODS AND
TOOLS TO SUPPORT GASTROINTESTINAL
OUTBREAK INVESTIGATION**



Adedoyin Abiola Awofisayo-Okuyelu

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

ACKNOWLEDGEMENTS

The writing and completion of this thesis was made possible by the contribution and efforts of several people, but first, I want to thank God for making this dream and aspiration become a reality. This research was funded by NIHR Health Protection research Unit in Gastrointestinal infections, and I acknowledge my funding body for this opportunity.

I am deeply indebted to my supervisors, Professor Noel McCarthy and Dr Ian Hall, for your guidance and supervision over the last three years of this research project. I appreciate your patience and your willingness to teach me, especially at the beginning of the research and when I ventured into the unknown world of mathematical modelling. I also appreciate your understanding for the time I needed to be away for an extended period, and the times I required a quick turn-around on manuscripts reviews. As a DPhil student in a research group with the opportunity to interact with other students and post-docs and observe the relationship with their supervisors, I know both of you went above and beyond for me to ensure I produced a high-quality piece of work. The long hours you spent reviewing my papers, cross-checking my analyses, confirming my equations and model outputs, cannot be repaid. You have made an invaluable contribution to my career, and all I can say is thank you, and promise to make you proud as I progress in my career path.

I acknowledge my collaborators on the different sections of my research: Prof Jeremy Hawker, Prof Bob Adak, Dr Julii Brainard, Ms Ifunanya Mgbakor, Dr Lisa Byrne, Dr Eve Arnold and Dr Adrian Pratt. I thank you for your comments and suggestions on my research. I acknowledge and thank my assessors, Dr Adrian Smith and Dr Jose Lourenco, for your contributions during my transfer of status and confirmation of status examinations. I thank Dr Babatunde Olowokure, for reviewing and proof-reading a couple of chapters.

I am extremely grateful to my wonderful family for your support and prayers. My amazing mother, Mrs Omolara Awofisayo, I appreciate your constant prayers, support and encouragement throughout the course of my program, I love you mum! My mother-in-law,

Mrs Okuyelu, for your prayers and always being concerned about my well-being, I say a big thank you. I couldn't have asked for a better 'mother-in-love'. My big brother and my lovely sister Adeolu and Eloho Awofisayo, thank you so much for your kind thoughts and constant words of encouragement. You always make me know how proud you are of me, and I am so glad I can share this achievement with you, I love you both. My very wonderful aunt, Mrs Itunu Olaniyan, and sister, Mrs Bukky Ajomale, thank you for looking after my stomach. You ensured that I had a constant supply of food so I didn't need to worry about cooking. Without your delicious meals, I doubt my brain would have functioned at its full potential. I appreciate and love you both.

My incredible husband, Femi Okuyelu, you deserve a chapter all to yourself, but unfortunately, I have to squeeze all of my appreciation into one paragraph. Words cannot begin to express my heart felt gratitude for everything you had to withstand during the course of my program. My cornerstone, my support system, my sounding board, my inspiration, my number one cheerleader, my super hero, my ATM machine and my best friend, we know what we have endured emotionally and financially over the last three years, and I thank you for supporting me and having my back through it all. Those days I called to vent out my frustration, the days when I thought I couldn't achieve this, the days I was just totally fed up; you were there for me, encouraging me and giving me tough love when I needed it. Indeed, this project would not be possible without you, and I am truly thankful. This is an achievement for both of us, and I know that the sky is just our stepping stone from this point. I love you babe, always and forever!

Finally, I remember you dad. This was our ambition, it was our goal from the start of my university education that I will go on to get a doctorate degree. I finally did it, and I just wish you were here to share in my joy, but I know you would have been very proud of me. This is for you dad!

I dedicate this thesis to Prince Richard Awofisayo and Pa Samuel Okuyelu. Rest on...

ABSTRACT

Incubation periods of common gastrointestinal pathogens are cited inconsistently without a solid evidence base for these distributions or factors affecting incubation time. In some high-profile outbreaks, such as the STEC O104 outbreak in Germany or the large 2009 STEC O157 outbreak in South East England, incubation periods were longer than expected complicating the collection of exposure data and the outbreak investigation process. This research aimed to use observational and experimental data to understand the incubation period distribution of some common gastrointestinal pathogens and identify factors that influence these distributions.

I conducted systematic reviews on the incubation period of *Campylobacter*, STEC and *Salmonella* Typhi. In addition to using traditional methods for testing and describing the pattern of heterogeneity, I developed methods for analysing data from observational studies which involved using Kolmogorov-Smirnov test and hierarchical cluster analysis. I also extracted and analysed individual patient data of cases associated with outbreaks of STEC reported in England and Wales. Finally, I developed a within-host mathematical model that simulated the process of *Salmonella* Typhi infection and conducted a sensitivity analysis to identify influencing parameters.

The systematic reviews and meta-analysis of individual patient data indicated the presence of heterogeneity across outbreaks which could not be explained by chance alone. Some factors associated with the distribution of incubation period were age, food vehicle, mode of transmission, attack rate and vaccine history. The mathematical model estimated an incubation period of 9.6 days for *Salmonella* Typhi, and parameters strongly correlated with incubation period included the net population of bacteria in the colon and bacterial growth during systemic infection.

Outputs of this research will inform the development of policies and public health interventions that contribute towards reducing the burden of disease due to gastrointestinal

infections. Further research, developing a stochastic model of the incubation period of *Salmonella* Typhi, is recommended.

PUBLICATIONS

A. Awofisayo-Okuyelu, I. Hall, G. Adak, J.I. Hawker, S. Abbott, N. McCarthy. A systematic review and meta-analysis on the incubation period of campylobacteriosis. *Epidemiol Infect.* 2017. (Published) (Chapter 4)

Awofisayo-Okuyelu AA, Brainard J, Hall I, McCarthy ND. Incubation period of Shiga-toxin producing *Escherichia coli*: a systematic review. *Epidemiologic reviews* (accepted for publication) (Chapter 5)

Awofisayo-Okuyelu A, Hall I, Arnold E, Byrne L, McCarthy N. Analysis of individual patient data to describe the incubation period distribution of STEC. *Epidemiology and Infections.* (responded to reviewers' comments) (Chapter 6)

A. Awofisayo-Okuyelu, N. McCarthy, I. Mgbakor, I. Hall. Incubation period of typhoidal salmonellosis: a systematic review and meta-analysis of outbreaks and experimental studies occurring over the last century. *BMC Infect Dis.* 2018;**18**:483. (Published) (Chapter 7)

Awofisayo-Okuyelu A, Hall I, McCarthy N. Pathophysiology of *Salmonella* Typhi as it relates to the distribution of incubation period. *Frontiers in Life Sciences.* (Submitted) (Chapter 8)

Awofisayo-Okuyelu A, Pratt A, McCarthy N, Hall I. Within-host mathematical modelling of the incubation period of *Salmonella* Typhi. *Royal Soc. Interface.* (Submitted) (Chapter 9)

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT.....	iii
PUBLICATIONS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES	xv
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS.....	xx
CHAPTER 1: INTRODUCTION.....	- 1 -
1.1 OVERVIEW OF GASTROINTESTINAL INFECTIONS	- 1 -
1.2 BURDEN OF DISEASE	- 2 -
1.3 SOURCES AND RISK FACTORS	- 4 -
1.4 INVESTIGATION OF GI INFECTIONS.....	- 6 -
1.5 INCUBATION PERIOD.....	- 6 -
1.6 AVAILABLE DATA FOR STUDYING INCUBATION PERIOD.....	- 8 -
1.6.1 Aggregated patient data from observational studies	- 8 -
1.6.2 Aggregated patient data from experimental studies	- 9 -
1.6.3 Individual patient data extracted from observational studies	- 10 -
1.6.4 Individual patient data extracted from experimental studies	- 10 -
1.7 MATHEMATICAL MODELLING OF HUMAN DISEASE DYNAMICS	- 10 -
1.8 CONCLUSION	- 12 -
CHAPTER 2: AIMS AND PROJECT PLAN.....	- 13 -
2.1 AIMS OF PROJECT	- 13 -

2.2	PROJECT PLAN & OVERVIEW OF METHODS	- 13 -
2.2.1	Systematic review and meta-regression	- 14 -
2.2.2	Meta-analysis of individual patient data	- 15 -
2.2.3	Mathematical modelling of incubation period	- 15 -
2.3	PRESENTATION OF RESULTS	- 16 -
CHAPTER 3: DEVELOPMENT OF A PROTOCOL FOR SYSTEMATIC REVIEW AND DATA SYNTHESIS OF OBSERVATIONAL STUDIES		- 18 -
ABSTRACT.....		- 18 -
3.1	INTRODUCTION.....	- 19 -
3.2	METHODS	- 21 -
3.2.1	Identifying optimal search strategy	- 21 -
3.2.2	Selection process.....	- 22 -
3.2.3	Quality assessment of data in identified studies	- 23 -
3.2.4	Data extraction	- 24 -
3.2.5	Statistical analyses	- 26 -
3.2.6	Test of statistical methods.....	- 30 -
3.2.7	Determining the distribution of incubation period	- 30 -
3.3	RESULTS.....	- 32 -
3.3.1	Identifying optimal search strategy	- 32 -
3.3.2	Bootstrapped vs Asymptotic p-values	- 34 -
3.3.3	Test of heterogeneity – I^2 and Kolmogorov-Smirnov test.....	- 36 -
3.3.4	Describing the patterns of heterogeneity - Hierarchical cluster analysis	- 37 -
3.3.5	Test of statistical methods.....	- 39 -
3.4	DISCUSSION	- 42 -
3.5	CONCLUSIONS	- 46 -

CHAPTER 4:	A SYSTEMATIC REVIEW AND META-ANALYSIS ON THE	
INCUBATION PERIOD OF CAMPYLOBACTERIOSIS		- 48 -
ABSTRACT.....		- 48 -
4.1	INTRODUCTION.....	- 49 -
4.2	METHODS	- 52 -
4.2.1	Research questions and modified PICO elements	- 52 -
4.2.2	Search strategy and selection process	- 52 -
4.2.3	Data extraction	- 53 -
4.2.4	Descriptive analyses.....	- 55 -
4.2.5	Statistical analyses	- 55 -
4.2.6	Subgroup analyses	- 57 -
4.2.7	Risk of bias	- 58 -
4.2.8	Determining the distribution of incubation period data	- 58 -
4.3	RESULTS.....	- 59 -
4.3.1	Literature search and selection process	- 59 -
4.3.2	Characteristics of studies included in the review	- 59 -
4.3.3	Test of heterogeneity.....	- 63 -
4.3.4	Factors that may explain heterogeneity.....	- 64 -
4.3.5	Identifying subgroups of studies using clustering analysis.....	- 64 -
4.3.6	Summary of subgroup analyses	- 68 -
4.3.7	Distribution of incubation period data.....	- 70 -
4.4	DISCUSSION.....	- 72 -
4.5	CONCLUSIONS.....	- 75 -
CHAPTER 5:	INCUBATION PERIOD OF SHIGA-TOXIN PRODUCING <i>ESCHERICHIA</i>	
<i>COLI</i> : A SYSTEMATIC REVIEW AND META-ANALYSIS.....		- 77 -
ABSTRACT.....		- 77 -

5.1	INTRODUCTION.....	- 78 -
5.2	METHODS	- 80 -
5.2.1	Research questions and modified PICO elements	- 80 -
5.2.2	Search strategy	- 81 -
5.2.3	Selection process.....	- 82 -
5.2.4	Data extraction	- 82 -
5.2.5	Descriptive analyses.....	- 84 -
5.2.6	Statistical analyses	- 84 -
5.2.7	Subgroup analyses	- 86 -
5.2.8	Risk of bias	- 86 -
5.2.9	Determining the distribution of incubation period data	- 86 -
5.3	RESULTS.....	- 87 -
5.3.1	Search strategy and selection process	- 87 -
5.3.2	Descriptive analyses.....	- 87 -
5.3.3	Test for heterogeneity.....	- 92 -
5.3.4	Factors that may explain heterogeneity	- 92 -
5.3.5	Identifying subgroups of studies	- 94 -
5.3.6	Summary of subgroup analysis	- 94 -
5.3.7	Distribution of incubation period data.....	- 95 -
5.4	DISCUSSION	- 98 -
5.5	CONCLUSIONS	- 101 -
CHAPTER 6: ANALYSIS OF INDIVIDUAL PATIENT DATA TO UNDERSTAND THE INCUBATION PERIOD DISTRIBUTION OF SHIGA-TOXIN PRODUCING <i>ESCHERICHIA</i> <i>COLI</i>		- 102 -
ABSTRACT.....		- 102 -
6.1	INTRODUCTION.....	- 103 -

6.2	METHODS	- 105 -
6.2.1	Case ascertainment.....	- 105 -
6.2.2	Comparing characteristics of excluded and included cases	- 105 -
6.2.3	Microbiological methods.....	- 106 -
6.2.4	Calculation of incubation period.....	- 106 -
6.2.5	Descriptive analysis and heterogeneity testing	- 106 -
6.2.6	Pattern of heterogeneity using hierarchical cluster analysis	- 107 -
6.2.7	Regression analyses	- 107 -
6.3	RESULTS.....	- 109 -
6.3.1	Comparing characteristics of excluded and included cases	- 110 -
6.3.2	Description of outbreaks.....	- 113 -
6.3.3	Heterogeneity between outbreaks.....	- 115 -
6.3.4	Pattern of heterogeneity	- 115 -
6.3.5	Description of cases	- 115 -
6.3.6	Reported symptoms.....	- 119 -
6.3.7	Incubation period and associated factors	- 120 -
6.4	DISCUSSION.....	- 126 -
6.5	CONCLUSIONS.....	- 129 -
CHAPTER 7: INCUBATION PERIOD OF TYPHOIDAL SALMONELLOSIS: A		
SYSTEMATIC REVIEW AND META-ANALYSIS		- 131 -
ABSTRACT.....		- 131 -
7.1	INTRODUCTION.....	- 133 -
7.2	METHODS	- 135 -
7.2.1	Research questions and modified PICO elements	- 135 -
7.2.2	Literature search	- 135 -
7.2.3	Selection process and quality assessment.....	- 136 -

7.2.4	Data extraction process.....	- 136 -
7.2.5	Descriptive analyses.....	- 138 -
7.2.6	Statistical analyses	- 138 -
7.2.7	Subgroup analysis	- 140 -
7.2.8	Risk of Bias.....	- 141 -
7.2.9	Determining the distribution of incubation period data	- 141 -
7.3	RESULTS.....	- 142 -
7.3.1	Literature search and selection process	- 142 -
7.3.2	Descriptive analysis of studies.....	- 143 -
7.3.3	Test for heterogeneity.....	- 150 -
7.3.4	Factors that may explain heterogeneity	- 150 -
7.3.5	Identifying subgroups of studies	- 152 -
7.3.6	Summary of subgroup analysis	- 153 -
7.3.7	Distribution of incubation period data.....	- 153 -
7.4	DISCUSSION	- 157 -
7.5	CONCLUSIONS	- 160 -
CHAPTER 8: REVIEW: PATHOPHYSIOLOGY OF <i>SALMONELLA</i> TYPHI AS IT RELATES TO THE DISTRIBUTION OF INCUBATION PERIOD.		- 161 -
	ABSTRACT.....	- 161 -
8.1	INTRODUCTION.....	- 162 -
8.2	METHODS	- 164 -
8.2.1	Literature search	- 164 -
8.2.2	Summarising the pathophysiology and extracting parameter values.....	- 164 -
8.2.3	Hierarchy of available evidence for parameter values.....	- 165 -
8.2.4	Selecting parameter values.....	- 165 -
8.3	RESULTS.....	- 167 -

8.3.1	Pathophysiology of Salmonella Typhi infection	- 167 -
8.3.2	Duration of the identified steps in the process of infection	- 167 -
8.4	DISCUSSION	- 180 -
8.5	PUBLIC HEALTH AND RESEARCH IMPLICATIONS	- 181 -
CHAPTER 9: WITHIN-HOST MATHEMATICAL MODELLING OF THE INCUBATION PERIOD OF <i>SALMONELLA</i> TYPHI		
	ABSTRACT	- 183 -
9.1	INTRODUCTION	- 184 -
9.2	MODEL FORMULATION	- 186 -
9.3	MODEL PARAMETERISATION	- 189 -
9.4	MODEL ANALYSIS	- 201 -
9.5	SOLVING THE MODEL AND INTERPRETATION	- 201 -
9.6	ESTIMATING INCUBATION PERIOD	- 204 -
9.7	EFFECT OF DOSE ON INCUBATION PERIOD	- 204 -
9.8	UNCERTAINTY SAMPLING AND SENSITIVITY ANALYSES	- 205 -
9.8.1	Uncertainty sampling	- 206 -
9.8.2	Sensitivity analysis	- 208 -
9.8.3	Uncertainty on the predicted incubation period	- 213 -
9.9	DISCUSSION	- 215 -
9.10	CONCLUSIONS	- 219 -
CHAPTER 10: DISCUSSION		
	DISTRIBUTION OF INCUBATION PERIOD DATA	- 223 -
	STRENGTHS AND WEAKNESSES OF METHODS AND DATA	- 224 -

PRACTICAL IMPLICATIONS OF STUDY FINDINGS	229 -
FURTHER RESEARCH RECOMMENDED.....	230 -
PERSONAL REFLECTION.....	232 -
REFERENCES	235 -
APPENDIX 1: SYSTEMATIC REVIEW OF THE INCUBATION PERIOD OF CAMPYLOBACTERIOSIS – PROTOCOL.....	257 -
APPENDIX 2: PROTOCOL FOR COLLECTION OF INDIVIDUAL PATIENT DATA -	266 -
APPENDIX 3: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF <i>CAMPYLOBACTER</i>	269 -
APPENDIX 4: DETAILS OF STUDIES INCLUDED AND DATA EXTRACTED IN THE SYSTEMATIC REVIEW AND META-REGRESSION OF THE INCUBATION PERIOD OF CAMPYLOBACTERIOSIS	275 -
APPENDIX 5: CULLEN AND FREY PLOT SHOWING POTENTIAL DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF <i>CAMPYLOBACTER</i>	276 -
APPENDIX 6: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF STEC.....	277 -
APPENDIX 7: DETAILS OF STUDIES INCLUDED AND DATA EXTRACTED IN THE SYSTEMATIC REVIEW AND META-REGRESSION OF THE INCUBATION PERIOD OF STEC	281 -
APPENDIX 8: CULLEN AND FREY PLOT SHOWING POTENTIAL DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF STEC	282 -
APPENDIX 9: STEC ENHANCED SURVEILLANCE QUESTIONNAIRE	283 -
APPENDIX 10: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF <i>SALMONELLA</i> TYPHI	289 -

APPENDIX 11: DETAILS OF STUDIES INCLUDED AND DATA EXTRACTED IN THE SYSTEMATIC REVIEW AND META-REGRESSION OF THE INCUBATION PERIOD OF <i>SALMONELLA</i> TYPHI	- 302 -
APPENDIX 12: CULLEN AND FREY PLOT SHOWING POTENTIAL DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF <i>SALMONELLA</i> TYPHI	- 303 -

LIST OF TABLES

Table 1 Checklist for assessing incubation period data reported by individual studies (adapted from Todkill et al)	24 -
Table 2 Details of data extracted from the studies	25 -
Table 3 Output and relevant articles identified from search terms used in PubMed using the pathogen <i>Campylobacter</i> as an example	34 -
Table 4 Summary mean of studies within subgroups from the clustering analysis of simulated data	40 -
Table 5 MeSH terms used in search strategy in the review of the incubation period of <i>Campylobacter</i>	54 -
Table 6 Characteristics of studies included in the review of the incubation period of <i>Campylobacter</i>	61 -
Table 7 Linear mixed effect and regression models showing effect of study characteristics on mean incubation period of campylobacteriosis.....	66 -
Table 8 Summary statistics of subgroups in the review of the incubation period of <i>Campylobacter</i>	68 -
Table 9 Characteristics of studies within subgroups in the review of the incubation period of <i>Campylobacter</i>	70 -
Table 10 MeSH terms used in search strategy in the review of STEC.....	81 -
Table 11 Details of information extracted in the review of STEC	83 -
Table 12 Characteristics of included outbreaks in the review of STEC.....	89 -
Table 13 Results of generalised linear regression model in the review of STEC	92 -
Table 14 Summary and characteristics of studies within subgroups in the review of STEC....	97 -
Table 15 Comparison of characteristics between excluded and included cases; and mean and median incubation period of cases included in the individual patient data analysis of STEC cases.....	110 -

Table 16 Characteristics of outbreaks included in the analysis in the individual patient data analysis of STEC cases	- 114 -
Table 17 Multi-level regression model showing factors associated with incubation period in the individual patient data analysis of STEC cases	- 122 -
Table 18 Details of data to be extracted in the review of S. Typhi	- 137 -
Table 19 Characteristics of included studies in the review of incubation period of S. Typhi....	- 144 -
Table 20 Additional characteristics of included studies in the review of incubation period of S. Typhi	- 147 -
Table 21 Linear regression analysis to identify factors associated with the distribution of incubation period in the review of incubation period of S. Typhi	- 151 -
Table 22 Summary statistics of outbreaks and incubation period within subgroups in the review of incubation period of S. Typhi.....	- 156 -
Table 23 Summary table of collated parameter estimates and type of evidence available for each step in the infection process	- 179 -
Table 24 Values, range and distribution of parameter values used in the sensitivity analysis. -	190 -
Table 25 Values from mathematical model and experimental study showing effect of dose on incubation period	- 205 -
Table 26 Partial rank correlation coefficients of model parameters	- 209 -
Table 27 Parameters obtained fitting Pearson Type 1 distribution to the incubation period data	- 224 -

LIST OF FIGURES

Figure 1 Schematic diagram of project plan	14 -
Figure 2 Venn diagrams showing search outcomes (left) and eligible articles (right) identified from all three databases	33 -
Figure 3 Scatterplot comparing regular p-values with bootstrapped p-values and the effect of sample size.....	35 -
Figure 4 Scatterplot showing difference in p-value against composite sample size	36 -
Figure 5 Dendrograms of average (left) and complete (right) methods showing dissimilarity matrix.....	38 -
Figure 6 Dendrograms of simulated datasets showing variation in incubation periods between studies.....	41 -
Figure 7 Flowchart of study selection process in the review of the incubation period of <i>Campylobacter</i>	60 -
Figure 8 Funnel plot testing small study effect in the review of the incubation period of <i>Campylobacter</i>	62 -
Figure 9 Collated epidemic curves re-created from raw data and arranged according to clustering analysis subgroups in the review of the incubation period of <i>Campylobacter</i> ..	63 -
Figure 10 Dendrogram showing compact visualization of dissimilarity matrix and identified subgroups in the review of the incubation period of <i>Campylobacter</i>	65 -
Figure 11 Forest plot showing mean incubation period and 95% CI in the review of the incubation period of <i>Campylobacter</i>	69 -
Figure 12 Goodness-of-fit plots for gamma distribution using maximum likelihood.....	71 -
Figure 13 Flowchart of study selection process in the review of STEC.....	88 -
Figure 14 Funnel plot showing relationship between sample size and mean incubation period in the review of STEC	91 -
Figure 15 Collated epidemic curves of included studies in the review of STEC.....	91 -

Figure 16 Dendrogram showing dissimilarity matrix and associated subgroups in the review of STEC	- 94 -
Figure 17 Forest plot showing mean incubation period and subgroup summary mean in the review of STEC	- 96 -
Figure 18 Flowchart showing selection of outbreaks and cases with available individual data in the individual patient data analysis of STEC cases	- 109 -
Figure 19 Forest plot showing median incubation period and IQR of outbreaks included in the individual patient data analysis of STEC case	- 117 -
Figure 20 Dendrograms showing the clustering pattern driven by the mode of transmission in the individual patient data analysis of STEC cases	- 118 -
Figure 21 Boxplot showing median days and inter-quartile range between onset of first symptom and reported symptoms in the individual patient data analysis of STEC cases	- 119 -
-	
Figure 22 Histogram of reported incubation period by mode of transmission and indicating cases with bloody diarrhoea in the individual patient data analysis of STEC cases.....	- 121 -
Figure 23 Flowchart of study selection process in the review of incubation period of S. Typhi	- 142 -
Figure 24 Funnel plot testing small study effect in the review of the incubation period of S. Typhi.....	- 145 -
Figure 25 Re-created epidemic curves of eleven studies in the review of the incubation period of S. Typhi.....	- 146 -
Figure 26 Dendrogram showing compact visualization of dissimilarity matrix and associated subgroups in the review of the incubation period of S. Typhi.....	- 153 -
Figure 27 Forest plot showing mean incubation period and subgroup summary mean in the review of the incubation period of S. Typhi.....	- 155 -
Figure 28 Hierarchy of evidence for parameter values derived from the literature in the review of the pathophysiology of S. Typhi	- 166 -

Figure 29 Theoretical flowchart showing steps of the infection process of Salmonella Typhi .	-
167	-
Figure 30 Flowchart of model showing the mathematical representation of the infection process.....	- 186 -
Figure 31 Data from Forest et al fitted to replication model in estimating the growth rate of bacteria in phagocytes with 95% confidence intervals.....	- 198 -
Figure 32 Model output showing the initial GI phase of the infection process in the within-host mathematical model	- 202 -
Figure 33 Model output showing the involvement of the CLn and phagocyte of the infection process in the within-host mathematical model.....	- 202 -
Figure 34 Model output showing bacteria in the CLn to secondary bacteraemia of the infection process in the within-host mathematical model.....	- 203 -
Figure 35 Probability distribution plots showing range and shape of parameter values used in the sensitivity analysis of the within-host mathematical model.....	- 207 -
Figure 36 Scatterplots showing the relationship of each parameter with the incubation period (parameters: delta to C)	- 210 -
Figure 37 Scatterplots showing the relationship of each parameter with the incubation period (parameters: rho to epsilon).....	- 211 -
Figure 38 Tornado plot showing direction and strength of correlation with incubation period .	-
212	-
Figure 39 Scatterplots showing correlation of parameters with incubation period in the sensitivity analysis	- 213 -
Figure 40 Empirical distribution function plot showing the predicted incubation period from 1000 simulated models	- 214 -

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
CDC	Centre for Disease Control and Prevention
CLn	Caecal Lymph node
DALYs	Disability Adjusted Life Years
ESQ	Enhanced Surveillance Questionnaire
GBRU	Gastrointestinal Bacterial Reference Unit
GBS	Guillain-Barré syndrome
GI	Gastrointestinal Infection
HC	Haemorrhagic Colitis
HUS	Haemolytic Uremic Syndrome
IID	Infectious Intestinal Disease
IP	Incubation Period
KS Test	Kolmogorov-Smirnov Test
LHS	Latin Hypercube Sampling
LMI	Low-Middle income countries
LRT	Likelihood Ratio Test
MLR	Multi-level Regression
MOI	Multiplicity of Infection
NESSS	National Enhanced STEC Surveillance System
PHE	Public Health England
PHEC	Public Health England Centre
PRCC	Partial Rank Correlation Coefficients
PT	Phage Type
RCT	Randomised Control Trial
SaSAT	Sampling and Sensitivity Analysis Tool

SCV	<i>Salmonella</i> Containing Vacuole
SSE	Sums of Squares
STEC	Shiga-toxin <i>Escherichia coli</i>
WHO	World Health Organization
WITS	Wild-type Isogenic Tagged Strain

CHAPTER 1: INTRODUCTION

1.1 OVERVIEW OF GASTROINTESTINAL INFECTIONS

Gastrointestinal (GI) infections, also known as infectious intestinal diseases (IIDs), are infections that cause gastroenteritis. Causative organisms could be viral (e.g. Norovirus, Rotavirus), bacterial (e.g. *Campylobacter*, *Salmonella*, *Escherichia coli*), or protozoal (e.g. *Giardia*, *Cryptosporidium*) (1). Infections could be acute or chronic, mild and self-limiting or severe (2). Typical symptoms of gastroenteritis include diarrhoea, vomiting and abdominal pain (1). Non-diarrhoeal GI infections such as typhoidal salmonellosis and listeriosis cause systemic infections and invasive diseases such as septicaemia and meningitis. Short and long term complications of gastroenteritis have been reported in subsets of patients, including those infected with self-limiting illness (3).

Preventing infection involves proper handwashing (4), adequate heating of food, avoiding cross-contamination and administering vaccine where available (5). In addition to implementing preventive measures, appropriate public health actions such as investigation of outbreaks and the recall of food or food products (6) may contribute towards the reduction of the burden of infection.

This thesis focuses on three common bacterial pathogens of GI infections: *Campylobacter*, Shiga-toxin *Escherichia coli* (STEC) and *Salmonella enterica* serovar Typhi (S. Typhi). These pathogens were selected because of their contribution to the overall burden of disease due to gastrointestinal infections.

1.2 BURDEN OF DISEASE

Gastrointestinal infections are a global public health issue resulting in a considerable morbidity and mortality, and are frequently reported. In 2015, GI infections were estimated to cause two billion cases of illness globally (7), of which 29% were foodborne. They resulted in over one million deaths (8) and 78.7 million disability-adjusted life years (DALYs) globally (9). In the United Kingdom (UK), approximately one in four of the population suffer from an episode of GI infection in a year resulting in a total of 17 million cases annually (10) and costing the National Health Service (NHS) an estimated £742.8 million (11). In Canada, there is an estimated 1.3 episodes of acute GI infection per year, and an estimated cost of \$3.7 billion (12).

Campylobacter

Campylobacter is the most common cause of bacterial gastroenteritis and the second most common cause of GI infections globally. Acute infections can result in serious long-term complications such as peripheral neuropathies, Guillain-Barré syndrome (GBS), reactive arthritis and irritable bowel syndrome (13,14).

The actual global incidence of campylobacteriosis is unknown due to incomplete data in low and middle-income countries (14), however, the incidence is said to have increased in the past decade (15). Even with poor data, campylobacteriosis is known to be endemic in low and middle-income (LMI) countries, and the annual incidence rates in industrialised countries is estimated to be between 4.4 and 9.3 per 1,000 population (14). In Europe, the incidence rates ranges from 29.9 to 13,500 per 100,000 persons and the lowest rates are found in Finland and Sweden (15). In the United Kingdom (UK), incidence rates are reported to be about 9.3 cases per 1,000 person-years in the community resulting in 500,000 cases and 80,000 general practitioner consultations (16). In Australia, the annual community incidence rate is estimated at 1,184 per 100,000 persons (17).

Children are more at risk of developing campylobacteriosis, especially in LMI countries, where the probability of infection decreases with age (15). In industrialised countries, the age-specific incidence rate of *Campylobacter* in children under 5 years is 43.4 cases per 100,000 person-years, and with a hospitalisation rate of 10.8%, it is the second most common cause of GI related hospitalizations (18). Associated case fatality rates have been estimated to range from <0.01% to 8.8% depending on the infected population (14). Approximately one-third of GBS cases globally have been attributed to *Campylobacter* infection and the annual incidence of reactive arthritis following campylobacteriosis is about 4.3 per 100,000 persons (14).

***Salmonella enterica* serovar Typhi (S. Typhi)**

Typhoid fever is an enteric infection caused by the bacteria *S. Typhi*. Between 1984 and 2000, the global incidence of typhoid fever increased by 31% from an estimated 16 million to 21 million cases annually (19). However, the associated case fatality rate decreased from 3% to 1% annually (19) and has also remained stable. Published reports in 2014 estimated the global incidence at 21 million cases (20) and an incidence rate of typhoid fever at 355 per 100,000 persons per year with an annual case fatality rate of 1.1% (21). Compared with the rest of the world, Europe and North America have the lowest incidence rates with 3 and <1 cases per 100,000 persons per year respectively (21). In England, Wales and Northern Ireland, an average of 219 laboratory confirmed cases are reported annually (22).

Typhoid fever is an important cause of acute febrile illness in crowded and low-income populations with poor sanitation and contaminated food and water (19). In LMI countries, an estimated 17.8 million cases are reported to occur each year, with Central Africa (23) and certain regions of south-central Asia and south-east Asia having the highest incidence rates of over 100 per 100,000 persons per year (21). Incidence rates are particularly high among children and adolescents (19), typically peaking in children 2 to 4 years of age (23). In sub-Saharan Africa, hospital-based studies show that non-Typhi serotypes are more commonly reported than *S. Typhi* (19).

Shiga-toxin *Escherichia coli* (STEC)

STEC is a type of *Escherichia coli* that causes human infection sometimes resulting in bloody diarrhoea or haemorrhagic colitis (HC). Although the incidence rate of STEC is lower than *Salmonella* and *Campylobacter* (17,24), the severity of the complications that result from infection make it a major public health issue. A serious complication of STEC infection is haemolytic uremic syndrome (HUS) (13). This disease is defined by three main clinical features including acute renal failure, thrombocytopenia and microangiopathic haemolytic anaemia (25).

STEC is estimated to cause over two million acute illnesses annually worldwide with an incidence rate of 43.1 per 100,000 person-years (26). About 50% of infections occur in children under 15 years of age (26). Similarly, the number of HUS cases that have resulted from STEC infections are estimated to be just under 4,000 cases with about 60% occurring in children under 15 years (26). STEC infections have resulted in over 200 deaths and about 50% of them occur in the elderly over 60 years of age (26).

1.3 SOURCES AND RISK FACTORS

Gastrointestinal infections generally occur following ingestion of contaminated food or water (foodborne), close contact with infected individuals (person-to-person transmission), direct or indirect animal contact, or contact with contaminated surface water (waterborne) (4,27).

Foodborne transmission can occur through consumption of land animals, plants or seafood (27) and accounts for 36% of all GI infections, of which 67% are viral, 30% are bacterial and 3% are parasitic (2). In the United States of America (USA), 51.1% of foodborne illnesses are attributable to plant-based foods, 41.7% to food products from land animals including poultry and red meat, and 6.1% to seafoods (28). In England and Wales, 30% of foodborne illnesses are attributed to consumption of poultry products, 17% to red meat and 5% to

plant-based foods (29). In Australia, eggs and egg products are associated with the highest burden of illness, followed by poultry products, vegetable and salad and seafoods (30). Non-foodborne routes of transmission include environmental exposure, animal contact and person-to-person transmission (31).

International travel and consuming restaurant meals are some of the risk factors that have been associated with campylobacteriosis (32,33). The commonly reported food vehicles include poultry (34) and meat products, which have higher risk of infection when prepared at a restaurant (32) as well as dairy products including raw milk (35). Raw seafood and raw oysters have also been implicated (32). Non-foodborne sources of *Campylobacter* include animal and farm exposure such as living on a farm, visiting a farm, contact with farm animals (36) or pet animals (32,35). Drinking untreated surface water or swallowing water from swimming pools are also associated with illness (32,37).

Similarly, international travel, use of antacids and consumption of restaurant meals are factors associated with human salmonellosis (38) as well as typhoid fever. In addition, pre-existing medical conditions and previous use of antimicrobials have also been identified as risk factors (38). In LMI countries, living in overcrowded households, consumption of raw vegetables (39) and outdoor toilet facilities (40) are associated with developing typhoid fever. Other risk factors of typhoid fever include infected food handlers, improper cooking and improper cooling of meals (41). Some of the food vehicles implicated in typhoid fever infection include raw milk, contaminated water and vegetables. However, the real source of infection in most cases is thought to be an infected food handler (42).

Some risk factors associated with developing STEC infections are working or camping in rural areas, exposure to animals, eating restaurant meals, and person to person transmission (43–45). Implicating food vehicles identified include undercooked beef, dairy, raw vegetables and sprouts (43,44).

1.4 INVESTIGATION OF GI INFECTIONS

GI infections can affect people sporadically, without any known cause or food vehicle; however, when there are two or more cases linked by source, place or time; or reported cases are higher than what is expected for that time and place, this will be called a foodborne outbreak (46). Outbreaks are identified either through reports of ill patients by food businesses or general practitioners / family doctors or through routine surveillance (6). Sometimes, outbreaks are reported by members of the public or the media (6).

Controlling and effectively managing an outbreak relies on prompt identification of the source of infection and targeting appropriate public health interventions (47). Outbreak investigations provide a natural environment to study the pathogen and help identify the source of infection in order to apply targeted control measures. They also provide an avenue to understand certain parameters about the pathogen, such as incubation period. However, adequate knowledge of the incubation period is necessary to undertake an effective outbreak investigation. This creates a cycle where understanding certain parameters aid the investigation and carrying out the investigation increases our knowledge of these parameters.

1.5 INCUBATION PERIOD

Incubation period of GI infections is the time from exposure to a contaminated source to onset of symptoms (48). Understanding the pattern of distribution of incubation periods offers insight into the characteristics of GI pathogens *in vivo* and enables predictions and inferences about the incubation period to be made in the future. Knowledge of incubation period is useful to understand the pathogenesis of the organism and develop mathematical models (of infection and transmission) to estimate the possible extent of the spread of infection (49).

During investigations of GI outbreaks, accurate incubation periods are necessary to:

- estimate the period of likely exposure for point source outbreaks (49,50),
- identify and exclude travel related cases (51,52),
- distinguish secondary cases from primary cases,
- formulate a hypothesis (53),
- identify probable cases in the absence of laboratory results (54),
- determine or predict the end of an outbreak (55),
- trace contacts to identify possible infected persons and limit the spread of infection especially for pathogens where the period of infectiousness precedes symptoms (54).

In some prominent national and international outbreaks like the E. coli O157 outbreak in the UK in 2009 (56) and the STEC O104 outbreak in Germany in 2011 (57), the incubation periods reported were longer than expected. This resulted in challenging epidemiological investigations particularly when collating relevant exposure data and identifying cases. The evidence available for cited incubation periods of GI pathogens is poor and inconsistent. Typically, the incubation periods are cited as a range or a median and the factors affecting this distribution are either poorly understood or unknown.

In a global review of *Campylobacter* infection (15), the reported incubation period ranged from one day to three days while in another review in Canada, Europe and the United States (52), the reported incubation period was four days or less. In an experimental study of *Campylobacter jejuni*, the mean incubation period was reported as 2.2 and 2.8 days (58). Another experimental study reported a median of 0.8 days and 2.9 days with a range of 0.7 to 1.1 days and 1.4 to 4.5 days respectively depending on the ingested dose (59). A large outbreak in Finland caused by drinking water reported an incubation period of five days (60), and two outbreaks associated with milk in the UK and USA reported ranges of two to eleven days and one to ten days respectively (61,62).

According to a WHO report (63), the incubation period of typhoid fever can range from three days to more than 60 days. In a review conducted by Waddington et al (64), a human challenge study of *Salmonella Typhi* reported an incubation period ranging from 4 to 56 days with a median of 7.5 days. In another challenge study (65), the incubation period ranged from 6.5 to 13 days with a median of 8 days. Outbreak-associated incubation periods for typhoid fever have included ranges of 3 to 29 days, 5 to 12 days, 12 to 16 days and 13 to 22 days (66).

Understanding the incubation period distribution and identifying factors that influence the distribution may explain the variability in reported data.

1.6 AVAILABLE DATA FOR STUDYING INCUBATION PERIOD

Understanding the incubation period of gastrointestinal diseases involves analysing different types of data containing different levels of information. Each type of data can be analysed using a different method and the outcome observed will be based on the level of information available for analysis. The types of data that may be employed in studying incubation period include:

- aggregated patient data from observational studies (67);
- aggregated patient data from experimental studies (68);
- individual patient data from observational studies (69);
- individual patient data from experimental studies (70).

1.6.1 Aggregated patient data from observational studies

Observational studies such as outbreak investigations provide a natural environment to study the incubation period of a disease. As infection occurs in a natural setting, host and

environmental factors that may influence incubation period can be highlighted. In order to ensure precision of the incubation period, point source outbreaks, where exact time of exposure can be identified, are studied. Where this is unavailable or scarce, pooling maximally available data and applying statistical approaches for uncertainty can be used to estimate the distribution of incubation period (71).

Conducting a systematic review of observational studies combines several outbreak investigation reports where necessary information can be extracted for analyses. The data available for extraction is aggregated at outbreak level and additional relevant information on individual patients are unavailable.

1.6.2 Aggregated patient data from experimental studies

Experimental studies may involve infecting voluntary participants with a specific pathogen in order to study the pathophysiology or immunology of the disease. These studies are not actually designed to study the incubation period; however, they provide an opportunity to study this and other characteristics such as dose response as participants are closely monitored from the time of exposure up to the onset of illness (65). Experimental studies are performed in a controlled environment and some factors such as age, gender and food vehicle are predetermined. While this may not leave room to study these factors and their effect on incubation period, other factors such as vaccine response and attack rates can be studied and the accuracy of the reported incubation period is reliable. As in observational studies, conducting a systematic review of experimental studies combines the results of several studies and extracted data can be aggregated at study level.

1.6.3 Individual patient data extracted from observational studies

Data from observational studies can also be available at individual patient level, however, to use this level of data, detailed information beyond what is provided in the outbreak report is sought. This involves reviewing investigated outbreaks of GI infections in order to collate data on each patient involved in the outbreak resulting in joint analysis across outbreaks. This level of information makes it possible to study patient factors such as underlying medical conditions and presenting symptoms that are not available in aggregated data. Furthermore, outbreak level data is also available to study factors like attack rate and mode of transmission. Another advantage to analysing this type of data is the ability to distinguish patient dependent factors from outbreak dependent factors.

1.6.4 Individual patient data extracted from experimental studies

Similarly, individual patient data can be extracted from experimental studies and characteristics of participants can be studied individually and also collectively under each experiment. In addition, where mathematical models have been developed to study incubation period, this data can be used to validate such models as certain host factors are pre-determined and fixed.

1.7 MATHEMATICAL MODELLING OF HUMAN DISEASE DYNAMICS

Mathematical modelling is a step-by-step process of translating real scenarios into the language of mathematics in order to describe or predict the dynamics of the event (72,73). The process of developing any model starts with identifying the research question and relevant facts about the research area (72,74). Based on the area of interest, the type of model to be used is determined, model parameters are identified and the model is built (72).

It is good practice to validate the model by checking its outputs against existing data sets and finally, predictions can be made on scientific insights (72,74).

From over 100 years ago, mathematical models have been used to study the dynamics of infectious diseases (75,76) and have advanced over the last fifty years into interdisciplinary fields of mathematics and epidemiology (74). The use of mathematical models to study bacteria and viruses became common in the early 1980s following work by May and Anderson (75). With mathematical models, we can use current information about a disease or pathogen to conclude or make inference on the dynamics of the disease process. More importantly, they can be used to quantify and clarify uncertainties around the disease process. Mathematical models have been used to study, understand or predict various aspects of infectious diseases including: the spread of disease within a host population (75), the spread of a pathogen within a host (77), the immune process of the human body (78), the outcome of an epidemic, the impact of control measures (76), the development of successful control strategies (79) and to inform public health policy (74). According to Keeling and Danon (76), “models can also be used theoretically, as explanatory tools clarifying the fundamental principles of transmission”.

Mathematical models have been used to estimate the incubation period of diseases like influenza (48), dengue fever (80), acquired immune deficiency syndrome (81) and anthrax (82) amongst many others. Developing models involve the use of available data, from human or animal experimental or observational studies, to inform model parameters (83) in order to simulate actual biological processes. This interplay between data and models is a fundamental relationship where statistical estimation of parameter values from epidemiological data make models good predictive tools (76).

Compartmental models are a form of mathematical modelling where the population under study is divided into compartments (84). Using this concept and employing the theoretical use of mathematical models, pathogen parameters such as incubation period (48,70,82)

and dose response (83) have been studied. The models are developed using differential equations to describe disease dynamics and infection process at different time steps. Infectious disease parameters are events that are described as occurring continuously in time, rather than as discrete time intervals, hence, each time step is defined in related quantity as the rate of change in the population in one compartment over time (72). The benefit of using compartmental models is that they can start as very simple models with two or three compartments and can then be expanded into more complex models.

Within-host mathematical models are models of in-host pathogen kinetics that explore the mechanistic interactions between host and pathogen (85). They have been used to study diseases such as malaria (86,87) and cholera (88) and also to predict long term outcome of public health interventions (89).

1.8 CONCLUSION

Campylobacter, STEC and *S. Typhi* are gastrointestinal infections with substantial burden of disease. Despite the importance of the knowledge of accurate incubation period, the reported incubation periods of these pathogens are inconsistent and supporting evidence unavailable. This research project aims to employ methods such as systematic reviews of observational and experimental studies, individual patient data meta-analysis and a within-host compartmental mathematical model to understand the incubation period distribution and identify influencing factors. Outputs of this project will contribute towards reducing the burden of disease by enhancing outbreak investigations, risk assessments, public health response and decision-making, as well as informing the development of public health policies and contributing to the body of knowledge and evidence-based practise.

CHAPTER 2: AIMS AND PROJECT PLAN

2.1 AIMS OF PROJECT

The aims of this project were to:

- 1) Understand and describe the distribution of incubation period of common gastrointestinal pathogens
- 2) Identify factors that may influence the distribution of incubation period
- 3) Develop methods that support the synthesis of outbreak data

2.2 PROJECT PLAN & OVERVIEW OF METHODS

In order to achieve the set aims, the project was divided into three phases, each employing a method to analyse data from either observational studies such as outbreak reports, or experimental studies (Figure 1). These methods included:

- Undertaking a systematic review and meta-regression of observational and experimental studies
- Carrying out a meta-analysis of individual patient data from observation studies
- Developing a within-host mathematical model to estimate incubation period using data from a human challenge study of *S. Typhi*.

The pathogens that will be studied in this project include: *Campylobacter*, *Salmonella Typhi* and Shiga-toxin *Escherichia Coli*. More detailed methods are given in each result chapter.

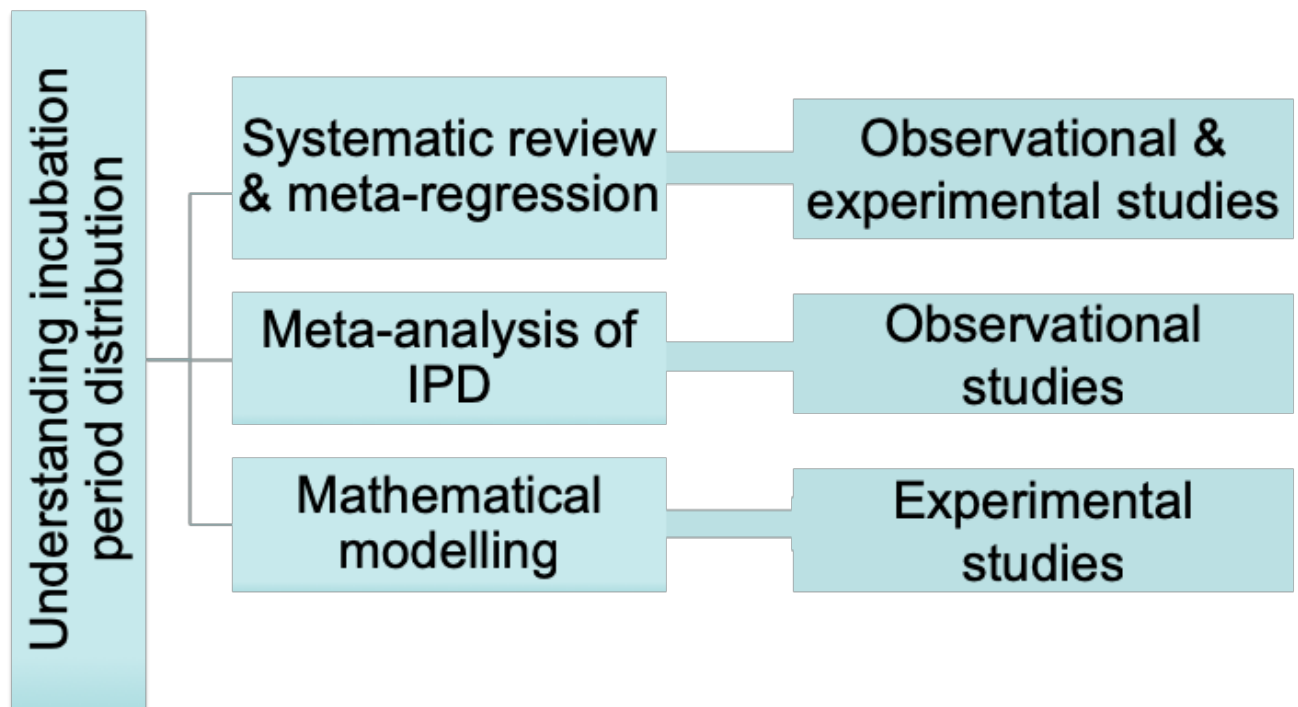


Figure 1 Schematic diagram of project plan

2.2.1 Systematic review and meta-regression

The first phase of the project involved undertaking a systematic review on the incubation period of each of the included pathogens. A protocol was developed for the systematic review of *Campylobacter* (Appendix 1) and then modified to suit the purpose of the systematic reviews of *Salmonella* Typhi and STEC. Published sources and grey literature of observational studies and experimental studies were reviewed for reported incubation periods of these pathogens. Data was extracted as either reported summary measures and/or individual incubation data. A meta-regression of the extracted summary and raw data was conducted in order to understand the distribution of incubation times and identify influencing factors.

As part of this phase, a protocol for synthesising data from observational studies (outbreak reports) was also developed. This will support the traditional methods of analysis and was developed simultaneously as the systematic review on the incubation period of

Campylobacter was underway. The developed protocol was then applied to data from the systematic review of *S. Typhi* and STEC.

2.2.2 Meta-analysis of individual patient data

The second phase of the project involved the extraction and collation of individual patient data (raw data) of outbreaks associated with STEC reported and investigated in England and Wales between 2009 and 2016. The data was retrieved from the Public Health England's national enhanced surveillance database of STEC, with the purpose of carrying out a meta-analysis on the collated data in order to identify factors affecting incubation time such as underlying conditions, current medications and previous infections.

A protocol was developed to describe the type and nature of data required, guide the data collection process and outline the data confidentiality requirements as requested by Public Health England (Appendix 2).

2.2.3 Mathematical modelling of incubation period

The third phase of the project focused on developing a within-host mathematical model that described the process of *S. Typhi* infection. A comprehensive literature review was undertaken to understand the pathophysiology of *S. Typhi* which was used to develop a mathematical flowchart of the infection process. A set of ordinary differential equations were developed and model parameters were identified from animal and human experimental studies. The developed model was validated and compared to experimental data from a human challenge study of typhoid fever.

2.3 PRESENTATION OF RESULTS

This section briefly describes the work conducted in this thesis and highlights the results that will be presented in the next seven chapters.

Chapter three reports the development of a protocol in synthesising data from observational studies. It states the justification for using additional approaches in conjunction with traditional methods in testing for heterogeneity and describing the pattern of heterogeneity. The method employed in conducting a literature review and the process of study selection are also described.

Chapter four reports the output of a systematic review of the incubation period of *Campylobacter* and the application of the developed methods. It outlines the process used in identifying and selecting included studies and the data extraction process. Using the developed methods, this chapter reports the observed heterogeneity and the clustering of studies into subgroups for further analysis. Factors influencing the distribution of incubation period are identified.

Chapter five describes the systematic review of the incubation period of STEC. As in chapter four, the literature review and study selection processes are outlined and the data extraction process is described. The application of the developed method is described and the results outlined. Influencing factors are identified and the estimated mean incubation periods across subgroups are reported.

Chapter six covers the methods and results of the meta-analysis of individual patient data extracted from reported outbreaks of STEC. The process of case ascertainment and criteria for the inclusion of cases is described. The regression model used to analyse the data and the methods used to identify similarities between cases and outbreaks is explained.

Strengths and limitations encountered with using individual patient data are discussed and recommendations are made for future work.

Chapter seven reports the systematic review of the incubation period of *Salmonella* Typhi. Similar to chapters four and five, the literature review and study selection process are outlined and the data extraction process is described. Also employing the developed methods, the observed heterogeneity is reported, as well as the identification of subgroups. Factors influencing the distribution of incubation period are identified.

Chapter eight presents the findings of a literature review on the pathophysiology of *S. Typhi* and how each step in the process of infection contributes to the incubation period. Data on the duration of each step are extracted, and where multiple studies report data on the same step, a selection is made on the most relevant data for research purposes. Justification for each selection is stated and public health implications based on the identified data is outlined.

Chapter nine describes the development of a within host mathematical model that simulates the infection process of *Salmonella* Typhi from ingestion to onset of clinical illness. A compartmental model that defines the process of infection using a set of ordinary differential equations is described. Model parameters were identified using animal and human experimental studies. The model outputs are presented and how this compares to the human challenge data of typhoid fever is discussed.

CHAPTER 3: DEVELOPMENT OF A PROTOCOL FOR SYSTEMATIC REVIEW AND DATA SYNTHESIS OF OBSERVATIONAL STUDIES

ABSTRACT

Systematic reviews of observational studies are increasingly being used especially in answering research questions that involve estimating the effect of interventions. While there are well developed guidelines on the process of undertaking such reviews, there aren't equivalent guidelines for synthesising data that measure parameters of infectious diseases.

This chapter describes the development of a protocol combining existing guidelines with additional methods in undertaking a systematic review to estimate the distribution of infectious disease incubation times from observational studies such as outbreak reports and experimental studies. The process of effectively searching and identifying relevant articles, the adapted criteria used in assessing studies and the extracting of different types of data is described. Additional methods for analysing available data in conjunction with traditional methods were employed. The methods of data synthesis employed included using I^2 test and Kolmogorov-Smirnov test (KS test) in testing for heterogeneity in the data reported, and hierarchical cluster analysis in defining subsets of studies with limited evidence of heterogeneity.

Possible limitations that may be encountered in analysing such data are highlighted and recommendations are made on the minimum data that should be available in outbreak reports and experimental studies to facilitate evidence synthesis by researchers on disease parameters using reported data.

3.1 INTRODUCTION

A systematic review aims to provide answers to well-defined research questions by collecting and summarising all available research evidence, and in doing so, is becoming more popular than traditional narrative reviews which provide a comprehensive background and identify existing gaps in research (90,91). Systematic reviews can be conducted on both experimental and observational studies and either qualitative or quantitative data can be synthesised. The process of conducting a systematic review involves framing a research question, identifying relevant studies, assessing the quality of studies, summarising or combining results if appropriate and interpreting findings or making recommendations (92). However, the methods used in each step have to be tailored to suit the research question being answered.

Observational studies are increasingly being used in research as there is a growing evidence of similarity of results with randomised control trials (RCTs) (93). Methods are well established for RCTs and have been adapted to analytical observational studies, however there is less standardisation in assembling and synthesising data from studies that report measurement of parameters; which are characteristics that define events. To my knowledge, there is no standardised method for testing heterogeneity and synthesising observational data that measures parameters that vary based on host-pathogen interaction, such as incubation period. In conducting a systematic review on incubation period, observational and experimental studies provide the necessary evidence, however, such experimental studies are rare. As available guidelines for undertaking such systematic reviews and synthesising data may not be completely applicable, additional methods may be required for assessing studies and testing for heterogeneity.

Knowledge of the incubation period of a pathogen is necessary to understand the host-pathogen dynamics. For pathogens where the period of infectiousness precedes symptoms,

accurate knowledge of the incubation period would help in contact tracing to identify possible infected persons and limit the spread of infection. It is also useful for effective outbreak investigations, developing mathematical models and contributes to the development of public health interventions.

While undertaking systematic reviews of observational and experimental studies to estimate the incubation period of common gastrointestinal (GI) pathogens, I developed and applied a protocol which combines existing guidelines with new approaches. Using this protocol, I will synthesise and analyse published evidence on pathogen incubation time, give an estimate of the mean incubation period, assess the shape of the distribution and identify factors that might influence the incubation period. In this chapter, I describe the development of the protocol using my review of *Campylobacter* as a reference, and the subsequent systematic review chapters describe the application of this protocol in synthesising the evidence on the incubation time of *Campylobacter*, *Salmonella* Typhi and STEC O157.

3.2 METHODS

My research questions were:

1. What is the optimal search mechanism to identify relevant articles?
2. What are the appropriate methods available for synthesising extracted data?
 - a. How can heterogeneity across outbreaks be tested?
 - b. In the presence of heterogeneity, how can studies be combined for analysis?

3.2.1 Identifying optimal search strategy

Many observational and experimental studies do not report incubation period as their primary results hence conducting a search for incubation periods without any criteria such as 'RCT' poses a difficulty in optimising specificity and sensitivity of search terms. This may result in many irrelevant articles having to be screened. Considering this, a pragmatic approach that focuses on the efficiency of search terms, and not just the completeness, needs to be developed.

I conducted several searches first with the aim of identifying the appropriate combination of search terms with optimised specificity and sensitivity and second, to determine the ideal databases for conducting the reviews.

3.2.1.1 Selecting the search terms

In conducting the literature search, I used a combination of words to identify a nested search term with a good balance between specificity and sensitivity. The selected search term will be sensitive, to include all potentially relevant articles and specific, to identify articles that largely focused on the area of interest. It will be expected to return the minimum number of articles from which a substantial number of relevant studies can be identified. In this phase of search term selection, the literature search was conducted in PubMed.

In the electronic database PubMed, I searched for the following words: “pathogen name” (e.g. *Campylobacter*), “humans”, “incubation”, “outbreaks” and “experimental”, combining them to create the following five search terms:

- Pathogen name (used alone)
- Pathogen name AND Humans
- Pathogen name AND Humans AND Outbreaks
- Pathogen name AND Humans AND Incubation
- Pathogen name AND Humans AND Experimental

In order to select a search term, the output of all five searches were screened and the proportion of articles that can be included in the review (number of eligible articles / total number of articles returned from search) was calculated. The search term with the highest proportion was selected.

3.2.1.2 Selecting among databases

The selected search term was then used in additional databases including Google Scholar and ISI Web of Knowledge to complete the literature search and assess the effectiveness of using additional databases. The outputs of the three databases were compared to identify unique articles in each database, as well as duplicate and triplicate articles. On completion of the screening process and quality assessment, the number of articles from each database eligible to be included in the review was also compared.

3.2.2 Selection process

Eligible articles were selected after the selection process was conducted in the following steps:

- Screening of titles and abstracts for articles with human infection.
 - Articles on any area of the pathogen other than human infection were immediately discarded and not reviewed further.

- Articles in any language other than English were also discarded.
- Screening of full text for reporting of summary or individual patient 'incubation period' data
- Review of full text to assess quality of data
- Further review of full text to assess exposure time and identify only definite point source outbreaks.

3.2.3 Quality assessment of data in identified studies

Articles resulting from searching all three databases were assessed for reporting of incubation period and quality of data. In most clinical trials, the primary goal of the study is estimation of the effect size, therefore, the quality assessment is designed to assess the quality of the studies in relation to their purpose (94,95). On the other hand, the quality assessment undertaken in this review focused on assessing the quality of the incubation period data reported and not the actual study. This was done because many of the studies did not necessarily set out to study incubation period, but rather to report on the process of an outbreak investigation or provide evidence on the source of infection in an outbreak. The set of criteria used in assessing the data was developed by a collaborator and used in a previous review (68). This method of quality assessment enabled me to effectively evaluate the quality of incubation period data reported and the accuracy of the estimation. The set of criteria and corresponding components are listed in Table 1 and a scoring system was used to assess the reported data. According to the scoring system, each component in the checklist was allocated one point, resulting in three points for each criterion, therefore the maximum number of points a study could be given was twelve. Studies scoring below six points were excluded as they did not meet the quality assessment criteria. Studies scoring between six and eight points were reviewed again by both reviewers to confirm that the reported incubation periods were reliable.

3.2.4 Data extraction

Data was extracted from the studies using a pre-determined format (Table 2). Summary data, including mean, median, mode and range, were extracted from all studies. Full data were extracted from studies that reported raw incubation data of individual cases as descriptive tables or epidemic curves. Where data was reported in the form of epidemic curves or graphical plots, I used an online data extraction tool called Web Plot Digitizer (96).

*Table 1 Checklist for assessing incubation period data reported by individual studies
(adapted from Todkill et al)*

Criteria	Component
Exposure	• Clearly defined exposure e.g. identification of implicated food vehicle or source patient • Exposure linked epidemiologically or microbiologically to outcome • Exclusion of other potential sources
Diagnosis	• Microbiological confirmation (human, food or environmental confirmation) • Specific and sensitive case definition for clinical cases • Time constraints on case definitions to exclude very early or very late cases
Accuracy of measurement	• Clearly defined exposure time (point source or continuous exposure) • Reliability of onset times considering method and delay of data collection during epidemiological investigation • Accuracy of reported onset time (hourly, 6-hourly, daily)

Ascertainment of bias	• Identification of exposed group and reporting of onset on all or part of exposed group • Exclusion of background cases • Exclusion of secondary cases and person to person transmission when studying an environmental or foodborne source

Table 2 Details of data extracted from the studies

Section	Information to be collected
General information	- Year of publication - Title of article - Authors - Type of publication (journals, conference abstract, grey literature, etc) - PubMed ID (where applicable)
Study characteristics	- Year of study - Study design (cohort, case-control, experimental, case series) - Country of study - Comments on method or quality of study
Pathogen characteristics	- Infectious agent - Species - Subtype
Outcome data/ results	- Case definition - Reported incubation period (individual data, mean, median mode and range)

	- Derived or calculated summary estimates incubation period (raw data extracted) - Source of calculated data (epidemic curve or author description) - Incubation period to particular symptoms
Factors that could affect incubation time	- No. of people affected / Total no. exposed (attack rate) - Age distribution - Setting - Mode of transmission - Food vehicle (for foodborne infections only) - Patient characteristics (e.g. previous infection or treatment, underlying illness)
Any other relevant information	- Any other relevant information

3.2.5 Statistical analyses

Following extraction of raw data, it was necessary to consider the consistency of data from the studies. Statistical analyses in a systematic review include:

- Testing for heterogeneity
- Describing the pattern of heterogeneity
- Meta-regression analysis

3.2.5.1 Testing for heterogeneity – I^2 and Kolmogorov-Smirnov test

In a traditional meta-analysis that compares the effect sizes, the I^2 is calculated by deriving the Cochran's Q-statistic and inputting it into the formula: $I^2 = 100\% \times \left(\frac{Q-df}{Q}\right)$; where Q is the Cochran's heterogeneity statistic and df is the corresponding degree of freedom (97). As I

collated the underlying distribution of individual level information which was reported as ordinal data rather than binary, it was not possible to derive the Cochran's Q-statistic. I therefore used an alternative test, Friedman test, which is the non-parametric alternative. It is used to test for differences between groups where the dependent variable is count data (98), and in this case, between the studies as the incubation period is count data. The output is similar to that of the Cochran's Q-statistic reporting the Chi-square distribution and the degree of freedom. These values were inputted into the above formula to calculate the value of I^2 .

A p-value of less than 0.05 from the Friedman test provided evidence of heterogeneity between the studies and the resulting I^2 value was interpreted according to the Cochran suggested threshold (99) to determine the magnitude of heterogeneity.

In addition to testing for differences between the studies, I wanted to compare the underlying distribution of the reported incubation period between studies, so I used the Kolmogorov-Smirnov test (KS test). The KS test is a non-parametric test performed to compare the cumulative distributions between groups, which gives a test statistic and a p-value (100). It measures variation in the position and shape of the distribution. In this analysis, the null hypothesis stated that data from two studies were sampled from populations with a single distribution. Because the reported incubation period data is recorded to the calendar day and the KS test cannot calculate exact p-values in the presence of ties in the datasets (or with large sample sizes), I also applied a bootstrapped version of the test with 10,000 samples to compare with the regular asymptotic approximation.

The bootstrapping method merges the two samples together and then draws from the aggregated sample (with replacement) two sub-samples of the size of the original split. The KS statistic is calculated for the new samples. This is repeated n-boot times and the KS statistic stored each for repeat. The original "real" KS statistic is then compared with the sampled distribution of KS statistics to derive a "p-value". If the aggregated sample is small

then the sampling with replacement may be misleading but for sufficiently large samples the bootstrapping method should be representative.

After selecting the appropriate p-value, I compared the selected p-values and employed the global test of statistical significance to confirm if any variation observed in the reported incubation period was due to chance. The specified significance level, α , was 0.05 for between sample comparisons with the KS test, and then I calculated the proportion of p-values that was less than this. If the observed variation was due to chance, the resulting proportion will be $\leq 5\%$. I calculated the probability of obtaining N p-values less than 0.05, and if less than <0.01 , then N is not due to chance.

3.2.5.2 Bootstrapped vs Asymptotic p-values

In order to confirm the most appropriate set of p-values to use for further analysis, I examined the effect of sample size on the regular and bootstrapped p-values. I combined sample sizes of both populations included in the pairwise test using a critical scaling that accounts for the sample size of the individual populations. The formula used was:

$\sqrt{n_1 n_2 / (n_1 + n_2)}$, where n_1 and n_2 are the sample sizes for population one and population two respectively, note that this formula is equivalent to the sample size scaling applied to the test statistic when calculating the approximate p-values to the Kolmogorov-Smirnov distribution.

3.2.5.3 Describing the patterns of heterogeneity - Hierarchical cluster analysis and Forest plot

Using the output of the KS tests as a measure of distance, hierarchical clustering provided a means to graphically visualise the dissimilarity between data reported by studies included in the review and also identify subgroups of studies based on limited evidence of heterogeneity (101).

In creating the hierarchical clusters, I subtracted the derived p-values from one to generate a dissimilarity matrix. The method of cluster identification was investigated between average linkage method or complete linkage method (102). In the average linkage method, the dissimilarities between the studies is calculated as the arithmetic mean of the distances between them (linking them by their average distance), while the complete linkage method chooses the maximum distance between the studies (linking them by their maximum distance).

In order to identify subgroups, a cut-off point for clustering needed to be determined. Considering the KS test was multiple pairwise test where multiple hypotheses were being tested, the probability of observing one significant result due to chance increases. Thus, the likelihood of incorrectly rejecting the null hypothesis, or making a type 1 error increases. I carried out a sensitivity analysis to investigate the impact of multiple testing by dividing the significance level (α) used in the KS test, 0.05, by 1 (no adjustments made), by the number of studies included in the test (partial adjustment), and by the number of hypotheses tested (complete adjustments). This was done to derive the adjusted critical significance level ($1-\alpha$) that can be used evaluate statistical support in the context of multiple testing.

As in a traditional meta-analysis, I also created a forest plot, but rather than showing the individual effect sizes and the overall effect, the forest plot showed the mean incubation period and the summary mean of subgroups identified from the hierarchical cluster analysis as well as the corresponding 95% confidence intervals.

3.2.5.4 Meta-regression analysis

Regression models were used to examine the association between outbreak characteristics and the incubation period. All outbreak characteristics available were included as explanatory variables in the models. A univariate model was used to examine the association of each explanatory variable with incubation period, and where statistical support

for association was observed, a multivariate model was built which included the associated variables.

Where sufficient individual incubation period data were extracted from the outbreaks, a linear mixed effect model was employed using transformed data to account for skewness. However, where this was not possible due to insufficient data, a generalised linear regression model was employed using the mean incubation period as the dependent variable. The distribution family used was gamma and the link function was identity.

3.2.6 Test of statistical methods

I applied the KS test and hierarchical cluster analysis to synthetic (dummy) data in order to test these methods and their application to describing incubation period data. I synthesized three sets of data using the distribution most suited to the real extracted data. All three datasets contained ten studies each sampled from populations with different means. The studies in the first dataset had sample sizes of five (dataset 1), similar to the data extracted from the systematic reviews, the second had sample sizes of ten (dataset 2), and the third had sample sizes of 100 (dataset 3), as a form of method validation. The data were simulated from a gamma distribution and for each study in the three datasets, the value of the scale parameter remained constant at one and the value of the mean was altered between one and nine days (note that the shape parameter is given by mean/scale), with an additional sample taken as the mean of the merged studies in the systematic review (5.3 days). The choice of distribution was based on evaluation of which distributions best described the incubation period data that I had extracted and collated. This is described in section 4.3.7 of chapter 4.

3.2.7 Determining the distribution of incubation period

In order to determine the underlying distribution of the incubation period data, raw data extracted from the studies were aggregated and the probability distribution that best fits the

data among a predefined family of distributions was specified. The family of distributions that were examined included normal, uniform, exponential, logistic, beta, lognormal and gamma.

The Cullen and Frey graph (skewness-kurtosis graph) was plotted to compare these distributions and identify the most likely distributions that best fit the data. As the sampling distribution of the skewness-kurtosis can be quite widespread, particularly with small sample sizes, a bootstrapped version of the test was applied with 500 samples.

The likely distributions were then fit to the extracted raw data by maximum likelihood (mle) to estimate distribution parameters and provide goodness-of-fit plots including the density, CDF, P-P and Q-Q plots.

3.3 RESULTS

3.3.1 Identifying optimal search strategy

3.3.1.1 Selecting the search term

When '*Campylobacter*' alone was used as a search term (search 1), it resulted in 15391 articles, and when the search terms 'humans' and '*Campylobacter*' were used (search 2), it resulted in 10038 articles. After excluding irrelevant titles and going through the screening and quality assessment processes, the proportion of eligible articles available for inclusion from search 1 and search 2 were 0.5% (73/15391) and 0.7% (66/10038) respectively (Table 3). The search terms in search 3, '*Campylobacter*' and 'humans' and 'outbreaks', returned 2747 articles, of which 1.9% (52/2747) were eligible for inclusion in the review. The proportion of eligible articles from searches 4 and 5 were 28% (7/25) and 1.4% (2/146) respectively (Table 3). Although these proportions were higher than other searches, the total number of articles returned were considerably fewer and the type of articles were very specific to *in-vitro* studies of *Campylobacter*.

The resulting proportion of eligible articles from search 3 and the total number of articles returned from the search, indicated the ideal balance between sensitivity and specificity. Approximately 12,500 and 7,000 fewer articles than searches 1 and 2 respectively were returned although 21 and 14 fewer articles respectively were included (Table 3).

Nevertheless, I selected search 3 as the most efficient and optimum search term that was specific enough to exclude the majority of irrelevant articles yet, including most articles from the two broader searches. It was the search term that returned the minimum number of articles and resulted in a satisfactory number of eligible articles.

However, studies with outbreak evidence made up a large proportion of the articles included in the review. Combining search 5 with search 3 to create a compound search string of '*Campylobacter*' and 'humans' and 'outbreaks' or 'experimental' will identify additional

human experimental studies, thereby making the output complete and the search process efficient.

3.3.1.2 Identifying the database

The compound search string was used in Google Scholar and ISI Web of Knowledge.

The search in Google scholar returned 41,400 results from which less than 1% (314/41,400) were selected for the screening process. A total of 301 articles were duplicates, already identified in other databases (250 in PubMed, 15 in ISI Web of Knowledge and 36 in both). Only 13 articles were unique to Google scholar alone (Figure 2).

ISI Web of Knowledge returned 1057 articles from which 59 (5.6%; 59/1057) were selected for full text screening. All of these were either duplicate or triplicate articles (15 in Google scholar, 8 in PubMed and 36 in both). There was no unique article identified from this database (Figure 2).

Following the selection process and quality assessment, a total of 42 articles were eligible for inclusion in the review, none of which were unique to any of the databases, and 17 articles were triplicates, identified in all three databases (Figure 2).

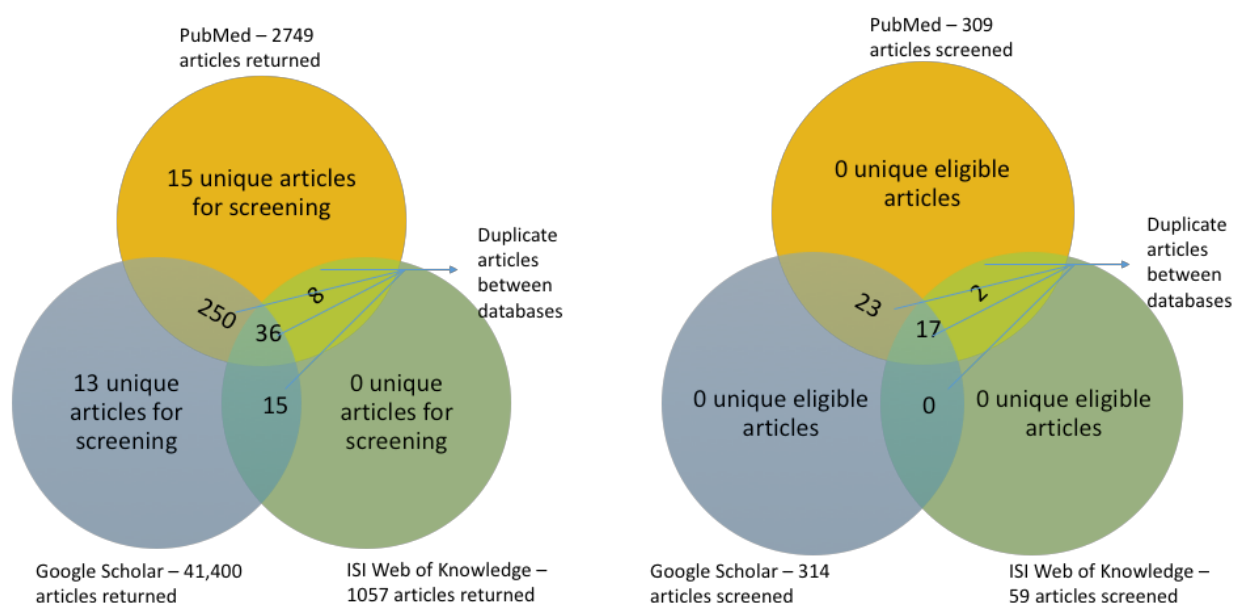


Figure 2 Venn diagrams showing search outcomes (left) and eligible articles (right) identified from all three databases

Table 3 Output and relevant articles identified from search terms used in PubMed using the pathogen *Campylobacter* as an example

	Search terms in PubMed	Number of articles returned from search	Number of articles included (%)	Additional articles to be screened to cover the difference
Search 1	<i>Campylobacter</i>	15,391	73 (0.47%)	0
Search 2	<i>Campylobacter</i> and humans	10,038	66 (0.67%)	5353
Search 3	<i>Campylobacter</i> and humans and outbreaks	2747	52 (1.89%)	12,644 or 7291
Search 4	<i>Campylobacter</i> and incubation and humans	25	7 (28)	
Search 5	<i>Campylobacter</i> and humans and experimental	146	2 (1.36)	

3.3.2 Bootstrapped vs Asymptotic p-values

I plotted the asymptotic p-values against the bootstrapped p-values to investigate the impact of the underlying data not being continuous. The scatterplot showed the bootstrapped p-values were more conservative than the asymptotic p-values. The composite sample size was colour graded with small sample sizes as dark blue and larger sample sizes as light blue. The colour graded data points were evenly distributed across the graph (Figure 3) indicating no pattern. I further tried to confirm if sample size had any effect where most difference was observed between the two p-values (the middle of the curve). I plotted the difference in the p-values against the composite sample size and fitted a spline curve (Figure 4) to check if there was any pattern indicating a change according to the difference

in p-values. According to the scatterplot, where the p-values were equal, the total sample size ranged from about 2.5 to 8.5, and when differences were observed, the total sample size appeared evenly distributed as illustrated by the spline curve in the graph (Figure 4). This shows that there was no effect of sample size on the outcome of the KS tests.

As there was no apparent effect of sample size on the p-values, I decided to use the bootstrapped p-values for further heterogeneity analysis. I selected the bootstrapped p-values since these don't inherently rely on underlying data continuity and so are more appropriate though potentially conservative.

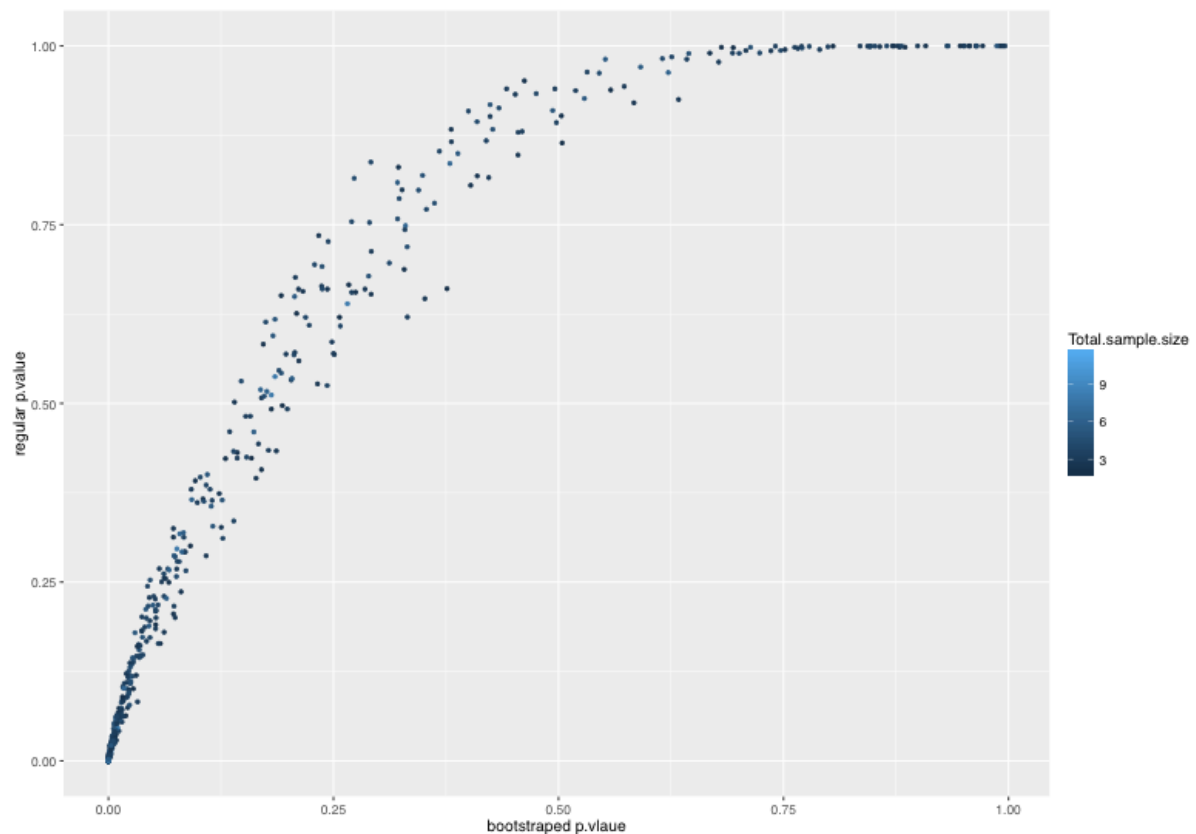


Figure 3 Scatterplot comparing regular p-values with bootstrapped p-values and the effect of sample size

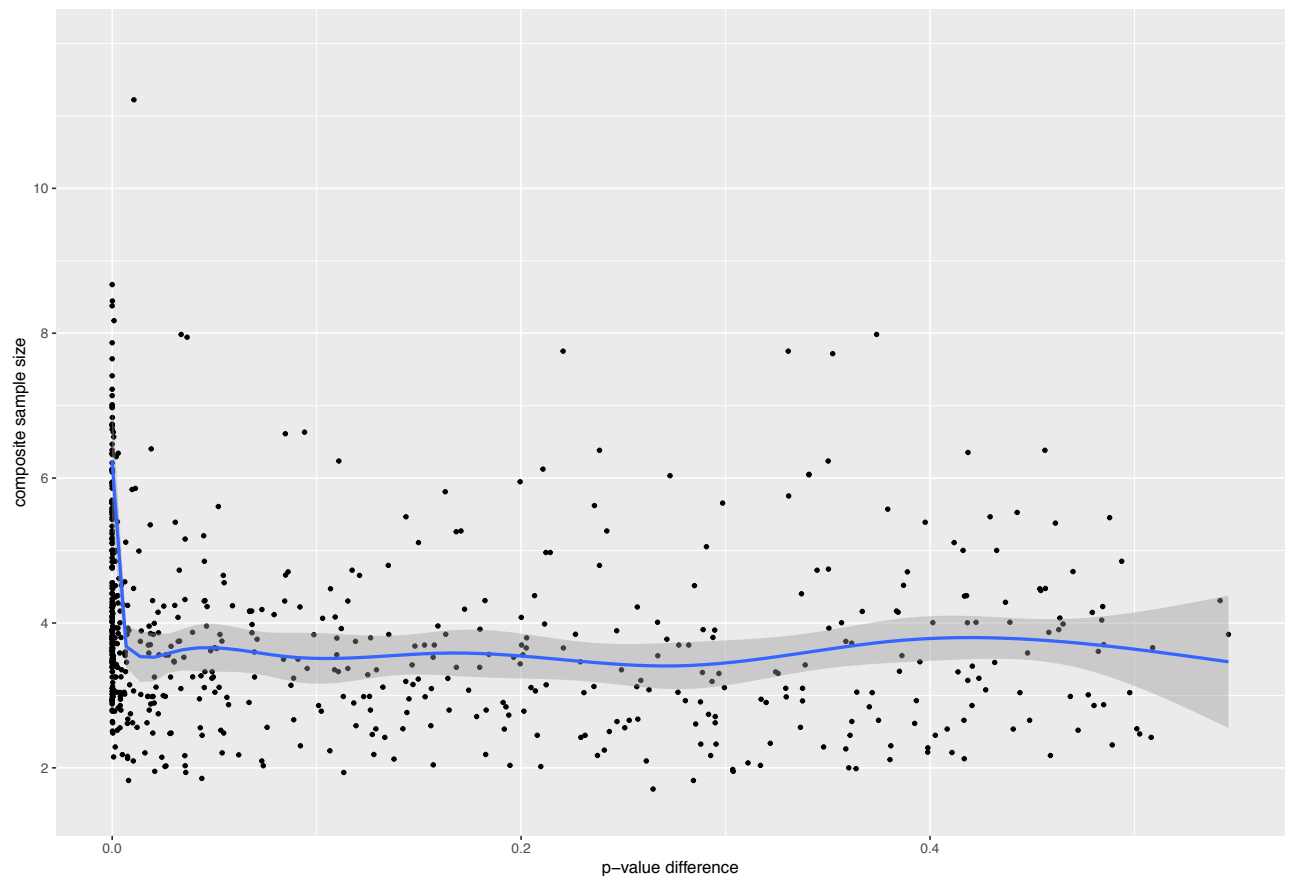


Figure 4 Scatterplot showing difference in p-value against composite sample size

3.3.3 Test of heterogeneity – I^2 and Kolmogorov-Smirnov test

The chi-squared value for the Friedman test was 104.3 with a degree of freedom of 29. The resulting p-value was <0.00001 showing evidence of heterogeneity between studies.

Inserting the output of the Friedman test into the I^2 formula, the value of I^2 was 72%.

According to the Cochran guide for interpretation (99), the resulting value indicates substantial heterogeneity.

Reviewing the KS test bootstrapped p-values, 5% were less than 0.05, and the probability of obtaining the resulting proportion was <0.00001 confirming the presence of heterogeneity which cannot be explained by chance.

3.3.4 Describing the patterns of heterogeneity - Hierarchical cluster analysis

The outputs were two dendrograms, showing compact visualisation of the dissimilarity matrix using first the average, then the complete methods (Figure 5). I observed identical pairing of some studies in both dendrograms, however, the magnitude of dissimilarity was different. Studies appeared more similar when using the average method and some studies that were paired in the complete method were unrelated in the average method indicating that the average distance was not exhaustive enough in defining the dissimilarities between the studies. The complete method preserved the maximum difference between the studies and thus, was more suitable for visualizing the differences.

Following the sensitivity analysis to determine a cut-off point for the subgroup specification, the resulting p-values were 0.05 (no adjustment), 0.0017 (partial adjustment) and 0.0001 (complete adjustment). The calculated cut-off points were 0.95, 0.9983 and 0.9999 respectively. The cut-off point of 0.95 was very selective, identifying more dissimilarities between the studies, however, the output could be subject to type 1 error as no adjustment was made to the p-value. The third cut-off point of 0.9999 had a stringent p-value that resulted in a really high and inclusive cut-off point which identified fewer dissimilarities between the studies and led to larger subgroups. The second cut-off point of 0.9983 that resulted from the partial adjustment was not as extreme as the first and not as conservative as the third. There was also a reduced possibility of type 1 error as a partial adjustment was made to the p-value. This cut-off point was an intermediate point as it was reasonably selective but also reliable, making it the appropriate cut-off point for classifying studies into subgroups.

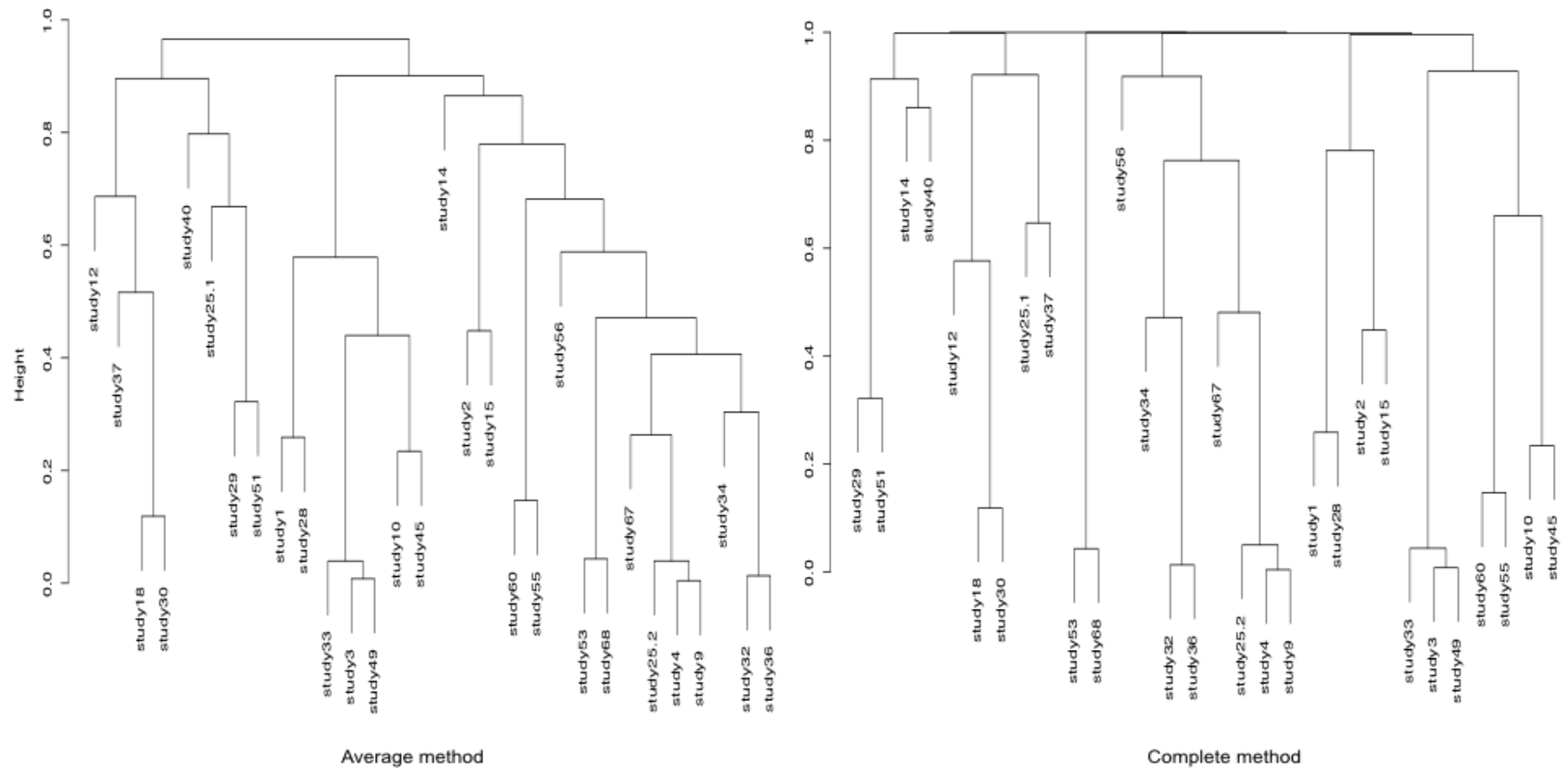


Figure 5 Dendrograms of average (left) and complete (right) methods showing dissimilarity matrix

3.3.5 Test of statistical methods

The sample mean of the studies in the simulated data for the three datasets ranged from 0.6 to 9.4 in dataset 1; 1.2 to 8.6 in dataset 2; and 1.0 to 9.3 in dataset 3; within the expected tolerances for sampling given the known population means.

In all three datasets, the proportions of KS test p-values below 0.05 were greater than 5% (dataset1: 26.7%; 12/45, dataset2: 64%; 29/45, dataset3: 95.6%; 43/45). The probability of obtaining the resulting proportions was <0.0001 , indicating high heterogeneity in the simulated data above what would be expected by chance, which is to be expected given the increasing population mean.

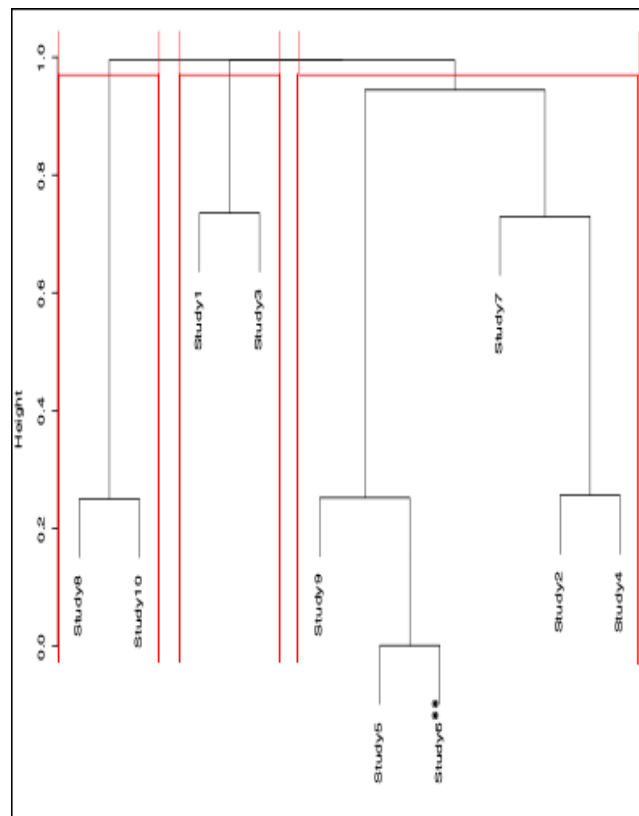
The data simulation and KS test analysis were repeated ten times in order to confirm the possible range of the proportion of p-values less than 0.05. Simulating a sample size of 5, the proportions ranged from 15.6% to 51.1% with an average of 32.4%. Simulating a sample size of 10, the proportions ranged from 55.5% to 81.8% with an average on 67.9%, and from the large sample size of 100, the proportion ranged from 86.7% to 95.6% with an average of 92.4%.

The hierarchical clustering analysis and the resulting dendrograms (Figure 6) show the amount of variation across the incubation periods in the ten studies. Dataset 1 resulted in three subgroups including one large subgroup of six studies and two small subgroups with two studies each. Dataset 2 also resulted in three subgroups containing five studies, three studies and two studies. Dataset 3 resulted in six small subgroups, one containing three studies, two containing two studies each and three subgroups of one study each.

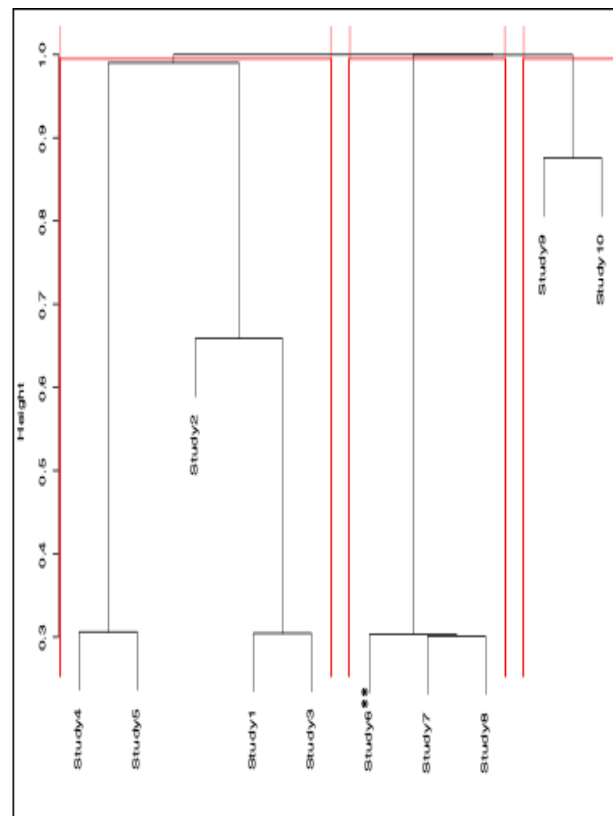
In all three dendrograms, studies with similar (sequential) means were clustered into the same subgroups. Studies with the smallest sample means were clustered into subgroup 1, and studies with the largest sample means were clustered into subgroup 3 (for datasets 1 and 2) and subgroup 6 (for dataset 3) (Table 4).

Table 4 Summary mean of studies within subgroups from the clustering analysis of simulated data

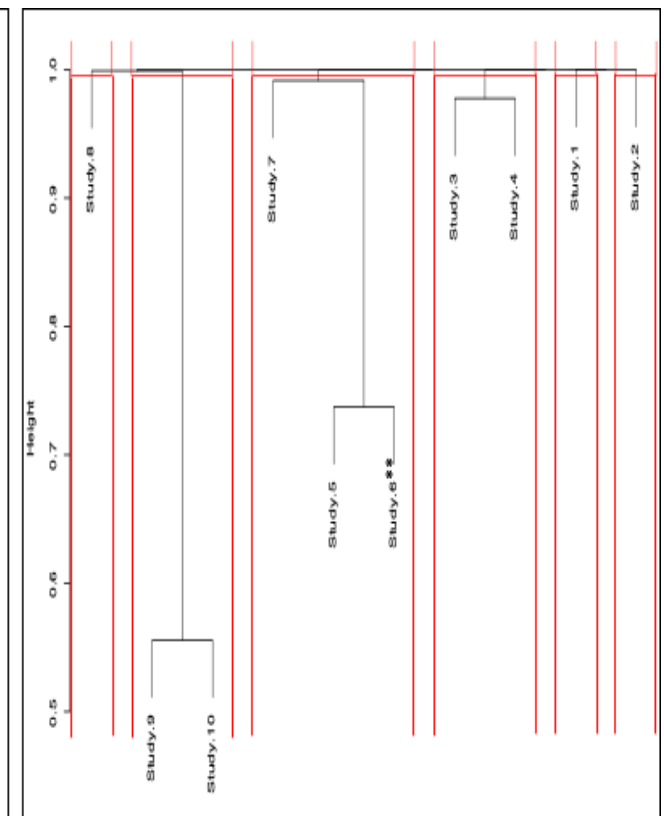
Subgroups	Dataset 1	Dataset 2	Dataset 3
	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Subgroup 1	1.8 (0.9 - 2.7)	3.2 (2.5 - 3.8)	1.0 (0.7 - 1.2)
Subgroup 2	5.4 (4.3 - 6.6)	6.6 (5.7 - 7.4)	1.9 (1.6 - 2.2)
Subgroup 3	7.1 (5.5 - 8.7)	8.0 (6.7 - 9.3)	3.7 (3.4 - 3.9)
Subgroup 4			5.3 (5.0 - 5.6)
Subgroup 5			7.5 (6.9 - 8.1)
Subgroup 6			8.5 (8.1 - 8.9)



dendrogram from dataset 1



dendrogram from dataset 2



dendrogram from dataset 3

** Mean of the merged studies in the systematic review (5.3 days)

Figure 6 Dendrograms of simulated datasets showing variation in incubation periods between studies

3.4 DISCUSSION

The process of systematic review and methods of meta-analyses of experimental studies, RCTs and observational studies have been reported (103–105), however, there are no specific guidelines on the review and synthesis of data from studies that involve measurements of parameters, most of which will be observational and many with other primary aim. Even with established guidelines, the process of systematic reviews needs to be tailored to suit the research question, likewise the method for identifying studies, assessing, extracting and analysing data on measurement of parameters needs to be adapted. Furthermore, these types of reviews do not measure the effect of a treatment or the outcome of an intervention, but rather examines a particular feature of a pathogen, hence the method used in analysing this data will be different from previously developed methods. Here, I have identified search terms to efficiently find articles reporting the parameter being measured, adapted the quality assessment criteria to ensure inclusion of articles with accurate measurements, identified different formats of available data (summary and raw data) and described the data extraction process, employed additional methods for testing heterogeneity and applied approaches to define subset of studies with no evidence of heterogeneity into subgroups.

Using a non-specific search term such as the pathogen name alone, although in theory is more sensitive, in practice, it may introduce the risk of misidentifying articles through human error due to the extremely high number of titles needed to be screened. Even with the recommended two or more reviewers screening titles to minimise error, this type of search is time consuming. There is now the possibility of conducting computer-aided systematic reviews (106), and where this is an option, a very sensitive search term is suggested.

The broad nature of searches required for this type of review results in a difficulty in attaining specificity of search terms. In developing search strings, it is advisable to use nested search

terms as this increases specificity. This is only suggested if there is sufficient data to allow a worthwhile meta-analysis and if there are time and cost constraints, however, where data is sparse, maximising sensitivity is advised. I found a combination of search terms that produced the relevant articles, however two different searches were undertaken to identify articles related to outbreaks and experimental studies.

It is advisable to search several sources of biomedical literature (107), however the effectiveness of Google Scholar can be debated, as a long list of irrelevant titles will need to be screened to identify few articles eligible for full text review. Furthermore, a substantial number of eligible articles can be identified in other databases.

Although ISI Web of Knowledge gives more specific results than Google Scholar, it includes more conference proceedings, which may not necessarily contain the level of information needed for measurement of parameters and as we observed, there were no unique articles identified in ISI Web of Knowledge as all articles were identified in other databases. Out of the three databases used in this review, I found PubMed to be a more effective database, and using it alone may be sufficient, particularly for rapid reviews.

Using the quality assessment checklist, I excluded studies where the measurement of exposure was unreliable. In these studies, incubation times were reported, but examining the studies in more detail showed that some authors may have inferred exposure time as the day of product purchase or distribution, which may not necessarily be the same as day of product consumption, which is when exposure happens (61). Also, some authors did not provide enough information or substantive evidence to identify the time of exposure, or to differentiate primary cases from secondary cases (108,109). The measurements used by these authors may have been useful for the purpose of their research, but for this study, they were not classified as a point source or single exposure. During the assessment phase, it is good practice to exclude studies without reliable information on precision of measurement to eliminate any uncertainty and improve the accuracy of the clustering analysis. It is also

recommended not to rely on the definition of point source exposures as given by the authors, but to review the characteristics of the outbreak in line with the assessment criteria.

If an assumption is made that the parameter measured follows a certain distribution, or if no variation is expected, then the summary data can be combined to estimate the outcome measure being studied. However, some variation would be expected as outbreaks/ exposures will vary (e.g. foodborne, airborne or person to person transmission) the size of the inoculum (shown to affect incubation period), vehicle of exposure (e.g. some foods are likely to protect against gastric acid) and susceptibility of those exposed. If variations are expected or observed, combining the raw data and carrying out a meta-analysis may be the appropriate way to estimate the parameter being measured. To test for heterogeneity, in addition to comparing the differences between the studies using the I^2 test, I compared the distribution of the reported incubation period data using KS test.

Other tests that could have been used for heterogeneity include the Mann-Whitney test and the Kruskal-Wallis test, both of which measure differences in the median of the distributions. The KS-test however uses the maximum difference between the curve, thereby testing the differences in position and shape of the whole distribution and not just one summary measure of the data. I therefore chose the KS test to allow comparison of the overall incubation time distribution rather than just a central measure. In traditional systematic reviews where outcomes of interventions are compared, heterogeneity can be further explored by testing the distribution of the magnitude of effect using the KS test or other test of distribution.

Having confirmed variation in the distribution of data reported by the studies, I needed to identify a suitable way to analyse the data. As suggested by Frumkin and Berlin(110), an ideal solution to combining studies with different exposures and measurements is to identify subgroups with similar characteristics. Forest plots and hierarchical cluster analysis provided the means to identify and group studies with similar characteristics and outcome measures.

A forest plot showed a graphical representation of the different mean incubation periods, hence, studies with similar summary measures could be identified. The hierarchical clustering was a method of identifying studies without evidence for heterogeneity and clustering them into subgroups. This was also a graphical representation showing the dissimilarities between the studies. I observed that the most appropriate method to accurately identify sub-groups of similar studies was the complete method using the bootstrapped p-values. I used the hierarchical clustering in conjunction with the forest plot by plotting studies according to the resulting subgroups identified by the clustering analysis. This validated the output of the clustering analysis, but also graphically magnified the differences between subgroups.

There was no obvious method to correct for multiple sampling conducted in the KS test, so I carried out pragmatic adjustments to the significance level and conducted a sensitivity analysis to ascertain the minimum and maximum number of subgroups that can be identified from the hierarchical cluster analysis. A Bonferroni correction was too conservative in identifying subgroups and an unadjusted p-value was too sensitive. In the absence of a clear way of deciding the actual cut-off point, I selected an intermediate point. This choice may not be generalizable and so further work is needed to rationalize choice when applied to other datasets and reviews.

Testing the methods used in identifying and describing the heterogeneity in the incubation period data (KS test and hierarchical cluster analysis) indicated that the statistical evidence for heterogeneity increased with sample size. Outbreaks with smaller sample sizes may appear similar on pairwise comparison, thus reducing the statistical evidence of heterogeneity between studies. Nevertheless, the global comparison is still able to show that heterogeneity, above chance expectation, is present.

A limitation of this approach is a lack of contextual information to identify a rationale for statistically clustered studies to be epidemiologically similar. However, such information is

also hard to extract from the studies reported, for example, ages were reported in broad categories like 1 – 79 years or summarised as adults and children. Without knowing what proportion of each age group was affected, it is difficult to estimate the effect of age on the incubation times and similarly for other case characteristics. Age groups should be described in functional categories that clearly differentiates adults from children, and the proportion of cases under each age band should be specified. Other host characteristics such as underlying illness or treatment and pathogen characteristics would also be useful if reported consistently.

It is recommended that a minimum requirement of data should be made available in outbreak and experimental reports if they are to be useful in data synthesis and epidemiological description of results to better understand pathogen incubation times. The protocol for literature search and data synthesis have been described and justifications for any selections made have been highlighted, however, as there are no guidelines available, some adjustments may be required to the developed methods when applying to other datasets.

3.5 CONCLUSIONS

I have identified in this chapter the optimum search terms and databases to be used in conducting an effective and rapid systematic review of observational studies reporting measurement of a disease parameter. I have also shown that employing KS tests and hierarchical cluster analysis, in conjunction I^2 test and forest plot, in the analysis of observational data from systematic reviews, particularly data that measure pathogen parameters, efficiently tests for the presence of heterogeneity and describes the pattern of heterogeneity. As these methods are unconventional in analysing systematic review data, some aspects of the method, such as selecting cut-off points, were done objectively based

on the data being analysed. Applying this protocol to the analysis of data from similar systematic reviews is highly recommended for validation.

CHAPTER 4: A SYSTEMATIC REVIEW AND META-ANALYSIS ON THE INCUBATION PERIOD OF CAMPYLOBACTERIOSIS

ABSTRACT

Accurate knowledge of pathogen incubation period is essential to inform public health interventions that contribute to the reduction of burden of disease. The incubation period distribution of campylobacteriosis is currently unknown with several sources reporting different times. Variation in the distribution could be expected due to host, transmission vehicle, and organism characteristics, however, the extent of this variation and influencing factors are unclear.

In this chapter, I have undertaken a systematic review of published literature of outbreak studies with well-defined point source exposures and human experimental studies to estimate the distribution of incubation period and also identify and explain the variation in the distribution between studies. I tested for heterogeneity using I^2 and Kolmogorov Smirnov test, regressed incubation period against possible explanatory factors, and used hierarchical clustering analysis to define subgroups of studies without evidence of heterogeneity.

The mean incubation period of subgroups ranged from 2.5 to 4.3 days. A variation was observed in the distribution of incubation period between studies that was not due to chance. A significant association between the mean incubation period and age distribution was observed with outbreaks involving only children reporting an incubation period of 1.29 days longer when compared with outbreaks involving other age groups.

4.1 INTRODUCTION

Campylobacteriosis is a zoonotic infection caused by a non-spore-forming Gram negative bacteria (111). The most common species reported in human diseases are *Campylobacter jejuni* (*C. jejuni*) and *Campylobacter coli* (*C. coli*) (112). In humans, the main route of transmission of *Campylobacter* is foodborne. Infection occurs following ingestion of undercooked meat and meat products as well as raw or contaminated milk and milk products. Infection can also follow contact with contaminated animals. Person-to-person transmission is rare but can happen. Abdominal cramps and diarrhoea are the most commonly reported symptoms. Non-specific symptoms that can also occur include headache, chills, fever and muscle pain. The duration of illness is usually about a week, with the severity declining after 24 to 48 hours, however 20% of cases may have a relapse (113,114).

According to the World Health Organization (WHO), *Campylobacter* sp. caused 96 million cases of foodborne illness worldwide in 2010 (9). It is the most commonly reported zoonosis in the European Union accounting for 45.2 cases per 100,000 people (115,116). In the United Kingdom (UK), there are approximately 9.3 undiagnosed cases in the community for every case reported to the national surveillance system (16), and an estimated 280,000 cases reported each year resulting in over 100 deaths (111,117).

A large proportion of reported cases are sporadic, however, outbreaks of campylobacteriosis have been reported with foodborne (118,119) and non-foodborne (120,121) sources identified. In the UK, 114 outbreaks were reported between 1992 and 2009, affecting a total of 2676 people (122). Outbreak investigation contributes to the reduction of the burden of disease by identifying the source of infection and informing public health strategies and policies. An effective outbreak investigation requires understanding of certain parameters of the infecting pathogen such as the expected incubation period distribution.

Incubation period, which is the time between exposure and onset of clinical symptoms, is also important for surveillance, risk assessment and effective public health response. In epidemiological studies, incubation period can be used to estimate the period of exposure, identify and exclude travel related cases, distinguish secondary cases and formulate a hypothesis (48). It can help in diagnosing possible cases in the absence of microbiological diagnosis (54) and also offers insights into clinical and public health practices (48). Essential to an outbreak investigation is constructing a case definition where a time restriction, sometimes based on the incubation period, is set to correctly classify cases as being part of the outbreak under investigation (6).

As a result of certain factors such as infectious dose, host factors and possibly, food matrix, the incubation period may vary between individuals. These, among other factors result in a distribution of incubation period. The incubation period distribution of campylobacteriosis is not clearly defined with different times being reported. The National Health Service in England and WHO report two to five days (123,124) while the Public Health Agency of Canada report one to ten days (125). Incorrect estimations may result in formulating inaccurate case definitions, wrongly defined exposure times, excluding outbreak cases as sporadic or travel related cases and vice versa (126) and misclassifying cases. It is therefore important to correctly estimate the incubation period distribution of campylobacteriosis to support effective outbreak investigations.

Point source outbreaks and human experimental studies, in which healthy volunteers may be infected with *Campylobacter* in order to study certain characteristics of the organism, provide an avenue to study the distribution of incubation period. Outbreaks are natural experiments and the outcome can be dependent on the effect of influencing factors, whereas, experimental studies occur in a controlled environment, with less unknown variation as a predetermined dose is administered, and characteristics of participants are screened to ensure similarities.

In this chapter, I systematically reviewed literature for outbreaks with well-defined point source exposures and human experimental studies. Reported individual patient incubation periods and summary estimates of the distribution of incubation period were extracted and analysed with the aim of describing the distribution of incubation period, identifying any variation in the distribution between outbreaks above expectation by chance, and attempting to explain any variation identified.

4.2 METHODS

4.2.1 Research questions and modified PICO elements

My research questions were:

1. What is the distribution of incubation period and the average (mean and median) incubation period of *Campylobacter* in humans?
2. Is there heterogeneity between the reported incubations times amongst studies?
 - a. Can any observed variation be explained?
 - b. What factors are affecting the distribution of incubation periods?

Population studied/Participants - Laboratory confirmed cases of *Campylobacter spp.* that form part of an outbreak or experimental infection.

Probable cases of *Campylobacter spp.* based on clinical symptoms and case definitions in the context of outbreaks

Infectious agent - *Campylobacter spp.* (all subspecies included)

Route of Infection - Foodborne and non-foodborne

Outcome - Onset of gastroenteritis as described or defined by the authors (diarrhoea, vomiting, nausea, abdominal cramps etc.)

4.2.2 Search strategy and selection process

A systematic literature search for peer reviewed publications of observational studies and experimental studies reporting incubation period was carried out on PubMed, Google Scholar and ISI Web of Knowledge. I searched for the following words: “*Campylobacter*”, “outbreaks”, “experimental”, and “humans”, combining common variations of the words to create search strings (Table 5). The reference lists of identified review papers were also screened to find other relevant studies where incubation period of *Campylobacter spp.* may

have been reported. The search was carried out between 21 January to 17 March 2016 and there was no restriction on the dates of articles returned or on the reported species. Articles in languages other than English were excluded.

Each article went through the selection and/or assessment stage which was done in the following phases:

- 1) Screening of titles and abstracts for articles with human campylobacteriosis
- 2) Screening of full text for reporting of incubation period data
- 3) Review of full text to assess quality of incubation period data reported.
- 4) Further review of full text to assess exposure times and identify outbreaks with confirmed point source exposures.

The quality assessment undertaken in this review focused on assessing the quality of the incubation period data reported based on a set of criteria and not the quality of the overall study (Table 1). This was done because many of the studies did not necessarily set out to study incubation period, but rather to report on the process of an outbreak investigation or provide evidence on the source of infection in an outbreak. Two reviewers were involved in the quality assessment stage, and where there was a difference in opinions, discussions were held until a consensus was reached.

4.2.3 Data extraction

Data was extracted from the studies using a pre-determined format (Table 2). General information on the published article, the study characteristics, as well as specific information on the outbreak or experiment, including attack rate and exposure, pathogen and patient characteristics which might influence incubation time, were extracted from each study according to the predetermined format. The outcome information to be measured was quantitative which was available as summary or raw data. All studies reported at least one summary statistic of the incubation period distribution as a mean, median, mode or range.

The unit of measurement was in days, and where this was reported in hours, I converted to days.

Some studies reported raw incubation period for individual cases either as an epidemic curve or a summary table. Where an epidemic curve was provided, the raw incubation period data was extracted using Web Plot Digitizer version 3.10, which is a free web-based data extraction tool (96). If a summary table was provided instead, the raw data was also extracted. Where both summary and raw data was provided, the raw data was used for analysis.

Table 5 MeSH terms used in search strategy in the review of the incubation period of Campylobacter

Search strategy:
Terms for
I. Campylobacter
II. Humans
III. Outbreaks
IV. Experimental
I. ("campylobacter"[MeSH Terms] OR "campylobacter"[All Fields]) AND II (I AND II)
II. ("humans"[MeSH Terms] OR "humans"[All Fields]) AND III (I AND II AND III)
III. ("epidemiology"[Subheading] OR "epidemiology"[All Fields] OR "outbreaks"[All Fields] OR "disease outbreaks"[MeSH Terms] OR ("disease"[All Fields] AND "outbreaks"[All Fields]) OR "disease outbreaks"[All Fields]) I AND II AND (I AND II AND IV)
IV. experimental [All Fields]

4.2.4 Descriptive analyses

Frequencies and percentages were calculated to summarise all studies according to the characteristics identified including: study design (observational or experimental), study type (cohort or case-control study), year of study, *Campylobacter* species, setting of outbreaks, age description of cases, mode of transmission and food vehicle, where applicable.

Using the extracted raw incubation data, histograms of reported incubation periods of individual cases were plotted to re-create the epidemic curves of the outbreaks. All epidemic curves were plotted using a uniform x-axis indicating the incubation period from zero to fifteen days and above, and an individual y-axis indicating the number of cases involved in each outbreak which varied according to the graph.

In order to confirm if the mean incubation period of the observational studies differ significantly from the mean incubation period of the experimental studies, I conducted a t-test.

4.2.5 Statistical analyses

The raw incubation period distributions extracted from relevant studies were used to test for heterogeneity in the reported data and describe the pattern of heterogeneity, while the summary statistics calculated from these and extracted summary statistics for outbreaks without individual patient data were used to identify factors that may explain heterogeneity. Statistical analyses were carried out using statistical software R version 3.2.3 (2015-12-10) – “Wooden Christmas Tree” (127).

4.2.5.1 Testing for heterogeneity

I tested for heterogeneity across studies by deriving the value of I^2 . A p-value of less than 0.05 from the Friedman’s test provided statistical evidence of heterogeneity and using the Cochran suggested threshold (99) the value of I^2 was interpreted to determine the magnitude of heterogeneity.

I also performed a two sample Kolmogorov-Smirnov test (KS test) to compare the cumulative distributions between the studies. I applied a bootstrapped version of the function with repeat sampling conducted 10,000 times in order to derive p-values that will provide improved coverage due to potential ties in the data comparisons. A small p-value indicated that the incubation period distributions were different, and the null hypothesis was rejected. I compared the resulting p-values to confirm if any variation observed was due to chance by calculating the proportion of p-values below 0.05. The probability of obtaining at least the observed proportion of p-values less than 0.05 was calculated, and if it was less than 0.01, this provided statistical evidence for variation in incubation time distribution.

4.2.5.2 Identifying factors that explain heterogeneity

In order to examine if the incubation period was influenced by the outbreak characteristics, I performed a linear mixed effect (random and fixed effects) analysis using the individual incubation period data provided as the dependent variable and the outbreak characteristics as the explanatory variables. I applied a square root transformation to the incubation period to reduce skewness of the data. Outbreak characteristics with sufficient information were included in a full multivariable model. Likelihood ratio tests was used as a means of attaining p-values by comparing the full model to an alternative model which excluded the variable of interest. A final model was developed by excluding variables without statistically significant association with incubation period ($p < 0.1$).

So as to allow the inclusion of studies reporting only summary data (mean), I further performed a linear regression analysis with the mean incubation period as the dependent variable. The effect of the explanatory variables on the mean incubation period was estimated by using a univariate model. Where statistical support for an association was observed ($p < 0.1$), a multivariate model was built which included the associated variables at that threshold to test for confounding.

Due to insufficient information, organism species was excluded as an explanatory variable in both analyses. The significance level for the final models was chosen to be 5%.

4.2.5.3 Identifying subgroups of studies for analysis

In the presence of statistically significant heterogeneity, I explored the data using subgroup analyses. However, rather than randomly allocating studies to subgroups, I employed hierarchical cluster analysis to identify subgroups of studies that can be combined. The bootstrapped KS test was used to create a hierarchical cluster to show a graphical representation of how the studies grouped together in terms of their dissimilarities. I subtracted the p-values from one to generate a dissimilarity matrix showing the distances between the samples. The cluster analysis algorithm used was the complete linkage method. The output was a dendrogram showing the compact visualisation of the dissimilarity matrix.

In order to reduce the likelihood of observing one significant result due to chance or making a type 1 error, I made pragmatic adjustments to the significance level (0.05) by dividing it by the number of studies included in the KS test which was 30. I then subtracted the adjusted p-value from 1 ($1 - \alpha$) to derive a cut-off point from which studies without evidence of heterogeneity can be defined within separate clusters. These clusters refer to subgroups of studies that do not have evidence of heterogeneity between them and can be combined for meta-analysis.

4.2.6 Subgroup analyses

I pooled the raw incubation data of studies within a subgroup to create a single dataset for each subgroup, and derived the following summary statistics:

- Number of studies included in a subgroup
- Total number of cases (sum of cases in all studies included in a subgroup)

- Mean and median incubation period of cases within a subgroup
- Standard deviation (SD), variance, skew and kurtosis of incubation period of cases within a subgroup

The mean attack rate of the studies within a subgroup was also calculated.

A forest plot showing the distribution of the mean incubation period and the corresponding 95% confidence interval was created. Studies without raw data (eight studies) were allocated to subgroups based on their reported mean and included in the forest plot, however, without a confidence interval as this could not be derived.

4.2.7 Risk of bias

I tested the data for 'small study-effect' using a funnel plot to visually examine the relationship between small sample sizes and incubation period.

4.2.8 Determining the distribution of incubation period data

The raw incubation period data extracted were aggregated, and a Cullen and Frey plot was first created to select the distributions that best fit the data. The data were then fitted to the selected distributions by maximum likelihood to estimate distribution parameters and the goodness-of-fit graphs and statistics were assessed.

4.3 RESULTS

4.3.1 Literature search and selection process

A total of 45,206 search results were retrieved from the three databases and the titles and abstracts were screened for relevance. Exclusion of articles considered irrelevant resulted in 682 articles, and after removing duplicates, 337 articles remained. An additional three articles were identified from searching through the reference list of review papers, resulting in 340 articles available for full text screening for incubation period data. Excluding articles that did not report incubation period and articles that did not meet the quality assessment criteria resulted in 60 articles remaining. These articles were further reviewed to ensure that the reported outbreaks were point source and the reported incubation period were accurate. Excluding outbreaks that were not point source (Appendix 3), 45 articles were included in the review (Figure 7). Four articles reported on two studies each bringing the number of studies included in the review up to 49 (Appendix 4). Of these, I was able to extract raw data from 30 studies.

4.3.2 Characteristics of studies included in the review

C. jejuni was the most commonly reported species accounting for 75.5% of included studies. Forty-five percent of the studies were published in year 2000 or later, and 81.6% were carried out in developed countries of Europe and North America (Table 6). Four studies were experimental and the remainder were epidemiological studies undertaken during outbreak investigations to identify the source of infection. Forty-six per cent of these (21/45) were retrospective cohort studies and 29% were descriptive studies.

The most common reported setting for outbreaks was private parties (14/49; 28.6%), including weddings and conference dinners, followed by farm visits (11/49; 22.4%). Poultry and dairy were the most frequently reported implicated food vehicle accounting for 40.8% (20/49) and 28.6% (14/49) respectively (Table 6). Comparing the food vehicle and setting of

the outbreak, 50% of outbreaks caused by poultry dishes occurred at a private party, and 57.1% of outbreaks caused by dairy or dairy products occurred during a farm visit.

The funnel plot created to test for small study-effect resulted in a symmetric funnel indicating that the size of the study did not have any effect on the reported incubation period (Figure 8). From the re-created epidemic curves, I observed a variation in the distribution of incubation period (Figure 9).

Based on the t-test, there was no significance difference between the mean incubation period of observational studies and experimental studies (p -value= 0.1).

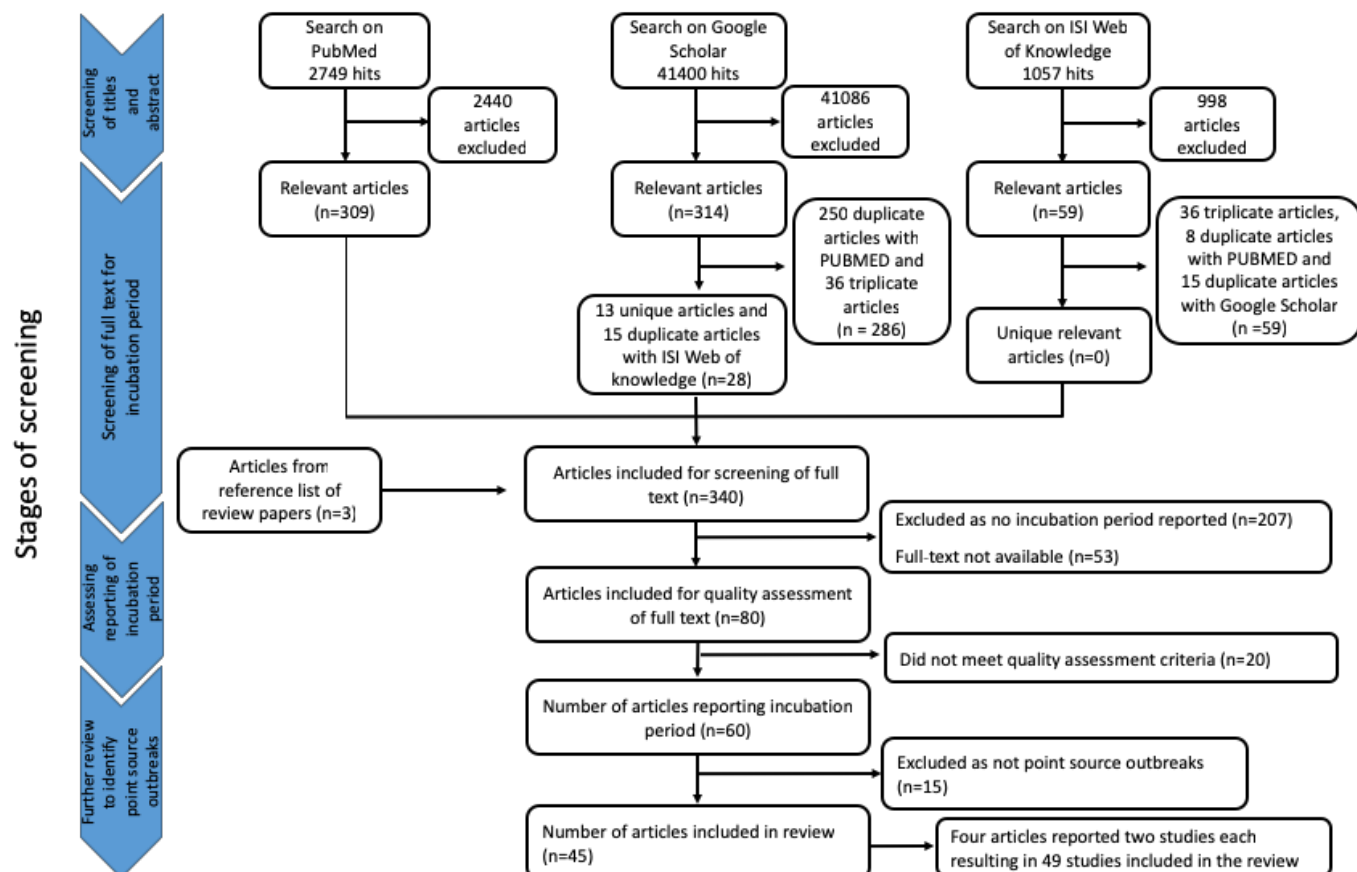


Figure 7 Flowchart of study selection process in the review of the incubation period of *Campylobacter*

Table 6 Characteristics of studies included in the review of the incubation period of *Campylobacter*

	N	%
Total number of studies	49	
Year of study		
Before year 2000	19	38.8
2000 and later	22	44.9
Unknown	8	16.3
Region of study		
Europe	20	40.8
North America	20	40.8
Australia	6	12.2
Asia	3	6.1
Species		
<i>Campylobacter jejuni</i>	37	75.5
<i>Campylobacter coli</i>	1	2.0
<i>C. jejuni</i> and <i>coli</i>	3	6.1
<i>C. jejuni</i> and <i>fetus</i>	1	2.0
Unknown	7	14.3
Age distribution		
Mixed ages	7	14.3
Children (0 – 18 years old)	15	30.6
Adult (over 18 years old)	27	55.1
Outbreak setting		
Private party	14	28.6
Farm visit / animal contact	11	22.4

Restaurants	10	20.4
Outdoor activity	5	10.2
School	5	10.2
Experimental study	4	8.2
Food vehicle category		
Poultry	20	40.8
Dairy	14	28.6
Water	1	2.0
Other	7	14.3
Unknown	7	14.3

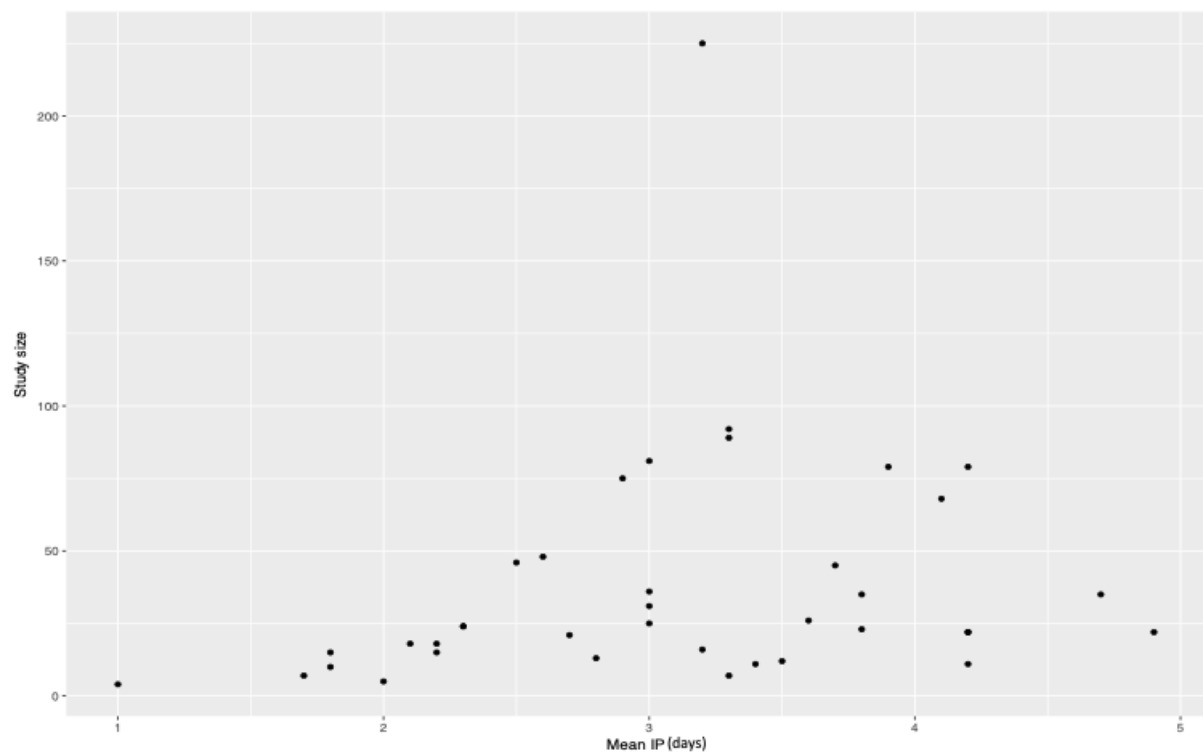


Figure 8 Funnel plot testing small study effect in the review of the incubation period of Campylobacter

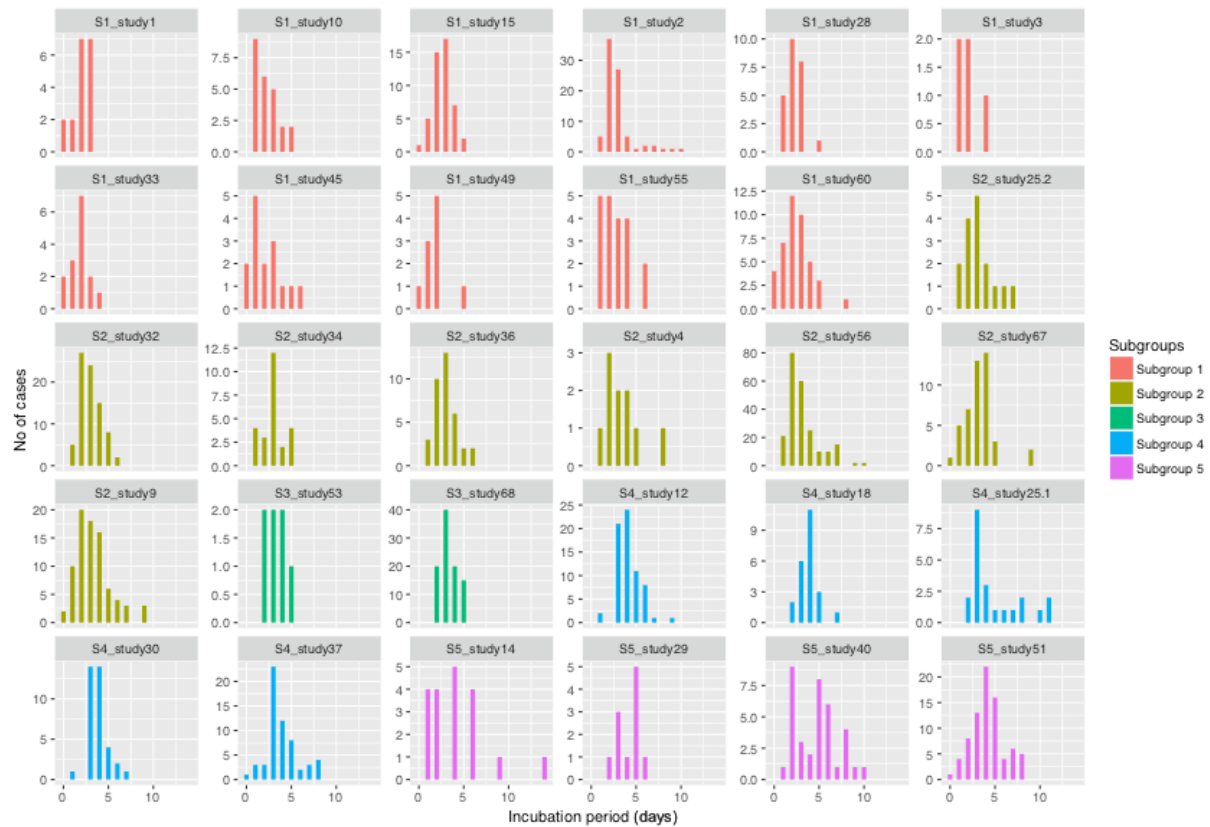


Figure 9 Collated epidemic curves re-created from raw data and arranged according to clustering analysis subgroups in the review of the incubation period of Campylobacter

4.3.3 Test of heterogeneity

The heterogeneity in the reporting of incubation periods across the different studies was $I^2 = 72\%$ (p-value for Friedman's test = <0.00001). The proportion of p-values from the KS test that was below 0.05 was greater than 5% (53%; 231/435). The probability of obtaining the resulting proportion was <0.00001 .

These results indicate a variation in the distribution of incubation periods between studies which is not due to chance alone.

4.3.4 Factors that may explain heterogeneity

From the linear mixed-effects multivariable analysis and the likelihood ratio tests, age distribution and outbreak setting were significantly associated with incubation period, while food vehicle category showed a weak association with a p-value of 0.08 and met the inclusion criteria into the final model (Table 7). Age distribution and outbreak setting remained significantly associated with incubation period ($p < 0.01$) in the final model after excluding the non-significant variables (attack rate and year of study) (Table 7).

From the linear regression univariate analysis, age distribution was the only variable with a significant association with the mean incubation period ($p < 0.01$) with outbreaks involving only children reporting a mean incubation period of 1.14 days longer when compared with mixed outbreaks involving both adults and children. In the final multi variable model also including outbreak setting, as one of the categories under outbreak setting met the inclusion criteria, the association with the mean incubation period remained significant ($p = 0.03$) with outbreaks involving only children reporting a mean incubation period of 1.29 days longer when compared with mixed outbreaks involving both adults and children (Table 7).

4.3.5 Identifying subgroups of studies using clustering analysis

Studies were paired and grouped based on evidence of dissimilarity. Studies found to have the least evidence of dissimilarity between them were paired. Likewise, some studies were not directly paired but attached to other pairs showing that the algorithm could not identify a single study with the least evidence of dissimilarity to them, but instead identified a pair of studies. The resulting output of this cluster analysis is presented as a dendrogram of the dissimilarity matrix (Figure 10).

Following the pragmatic adjustments made to the significance level, the resulting p-value was 0.0017 and the derived cut-off point was 0.9983. Five subgroups were identified using the cut point of 0.9983 to implement the p-value cut point of 0.0017, taking multiple testing

into account. These comprised: a subgroup of eleven studies, a subgroup of eight studies and three subgroups of five, four and two studies. (Figure 10).

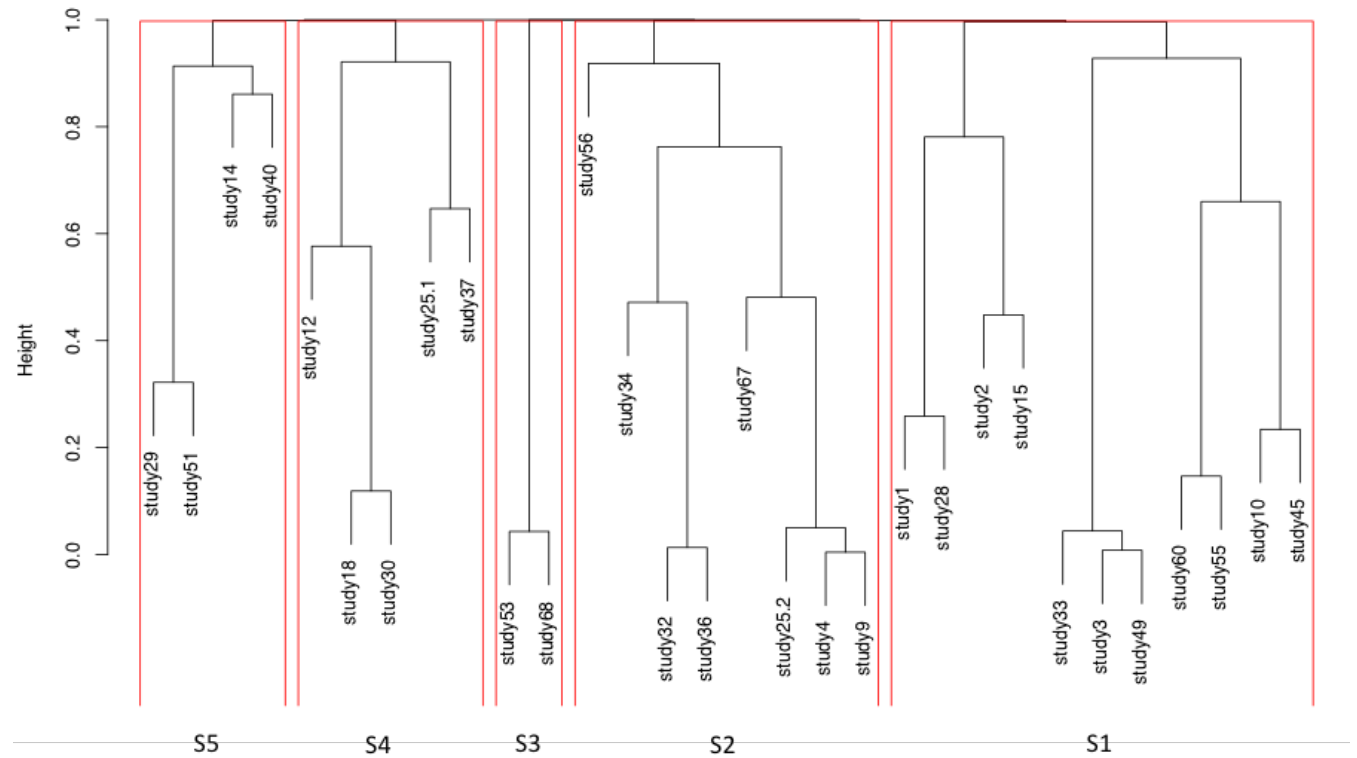


Figure 10 Dendrogram showing compact visualization of dissimilarity matrix and identified subgroups in the review of the incubation period of Campylobacter

Table 7 Linear mixed effect and regression models showing effect of study characteristics on mean incubation period of campylobacteriosis

Using raw data					Using summary data			
	Linear mixed effect full model		Linear mixed effect final model		Linear regression univariate analysis		Linear regression multivariable analysis	
Characteristics	Difference in mean incubation period	p-value of LRT	Difference in mean incubation period	p-value of LRT	Difference in mean incubation period	p-value	Difference in mean incubation period	p-value
Attack rate	0.007 (-0.002, 0.022)	0.10			-0.003 (-0.01, 0.01)	0.60		
Year of study		0.60						
After 2000	Reference				Reference			
Pre 2000	-0.1 (-0.3, 0.5)				0.1 (-0.4, 0.8)	0.57		
Age distribution		<0.001		0.005				
Mixed ages	Reference		Reference		Reference		Reference	
Adults	0.2 (-0.2, 1.1)		0.3 (-0.2, 1.2)		0.3 (-0.5, 1.1)	0.45	0.1 (-0.7, 0.9)	0.84
Children	1.7 (0.6, 3.3)		1.5 (0.5, 3.0)		1.1 (0.2, 2.0)	0.01	1.3 (0.1, 2.5)	0.03
Outbreak setting		0.01		0.001				
Other	Reference				Reference		Reference	
Farm visit	-1.2 (-1.2, -0.6)		-1.3 (-1.5, -0.6)		0.3 (-0.5, 1.1)	0.47	-0.4 (-1.5, 0.6)	0.41
Private party	-0.5 (-0.7, 0.1)		-0.5 (-0.9, 0.9)		-0.09 (-0.9, 0.7)	0.80	0.1 (-0.7, 0.8)	0.89

Restaurant	-0.7 (-0.9, -0.01)		-0.7 (-1.1, 0.03)		-0.8 (-1.7, 0.1)	0.08	-0.6 (-1.5, 0.2)	0.15
School	-1.1 (-1.1, -0.4)		-1.1 (-1.3, -0.3)		-0.4 (-0.5, 1.4)	0.37	-0.6 (-1.9, 0.7)	0.34
Food vehicle category		0.08		0.06				
Other	Reference				Reference			
Dairy	1.1 (-0.01, 3.2)		1.4 (0.2, 3.5)		-0.03 (-1.1, 1.1)	0.95		
Poultry	0.6 (-0.09, 1.8)		0.5 (-0.2, 1.5)		-0.41 (-1.5, 0.6)	0.45		

4.3.6 Summary of subgroup analyses

The subgroup containing eleven studies included 302 cases while the subgroup containing eight studies included 520 cases. The smallest subgroup with two studies also consisted of the lowest number of cases with 102 cases. The mean incubation period of studies in the subgroups varied between 2.5 days and 4.3 days (Table 8). There were also substantial differences in the variance, skew and kurtosis between subgroups (Table 8). There was some variation between the studies within subgroups (Figure 11) albeit not sufficient to evidence difference statistically.

The characteristics of four subgroups were quite similar in terms of the age distribution of cases and food vehicle (Table 9). These four subgroups included outbreaks which mostly reported poultry as the implicating food vehicle and at least 50% of the outbreaks involved only adults. Food services were reported as an outbreak setting in studies in four subgroups, however it was the predominant outbreak setting in subgroup 1. The characteristics of subgroup 4 were different with 80% of outbreaks involving only children; dairy products and farm were the most commonly reported food vehicle and outbreak setting respectively.

Table 8 Summary statistics of subgroups in the review of the incubation period of Campylobacter

	Freq.	Sum of	Mean	Median	Mean (95%	Variance	Skew	Kurtosis
		cases	attack		CI)			
			rate					
Subgroup 1	11	302	45.1	2	2.5 (2.3 – 2.7)	2.1	1.5	4.6
Subgroup 2	8	520	44.4	3	3.2 (3.1 – 3.4)	2.5	1.3	2.2
Subgroup 3	2	102	26.4	3	3.3 (3.1 – 3.5)	1.0	0.3	-0.9
Subgroup 4	5	208	51.3	4	4.1 (3.9 – 4.3)	2.7	1.4	3.3
Subgroup 5	4	145	46.4	4	4.3 (3.9 – 4.7)	4.7	0.8	2.0

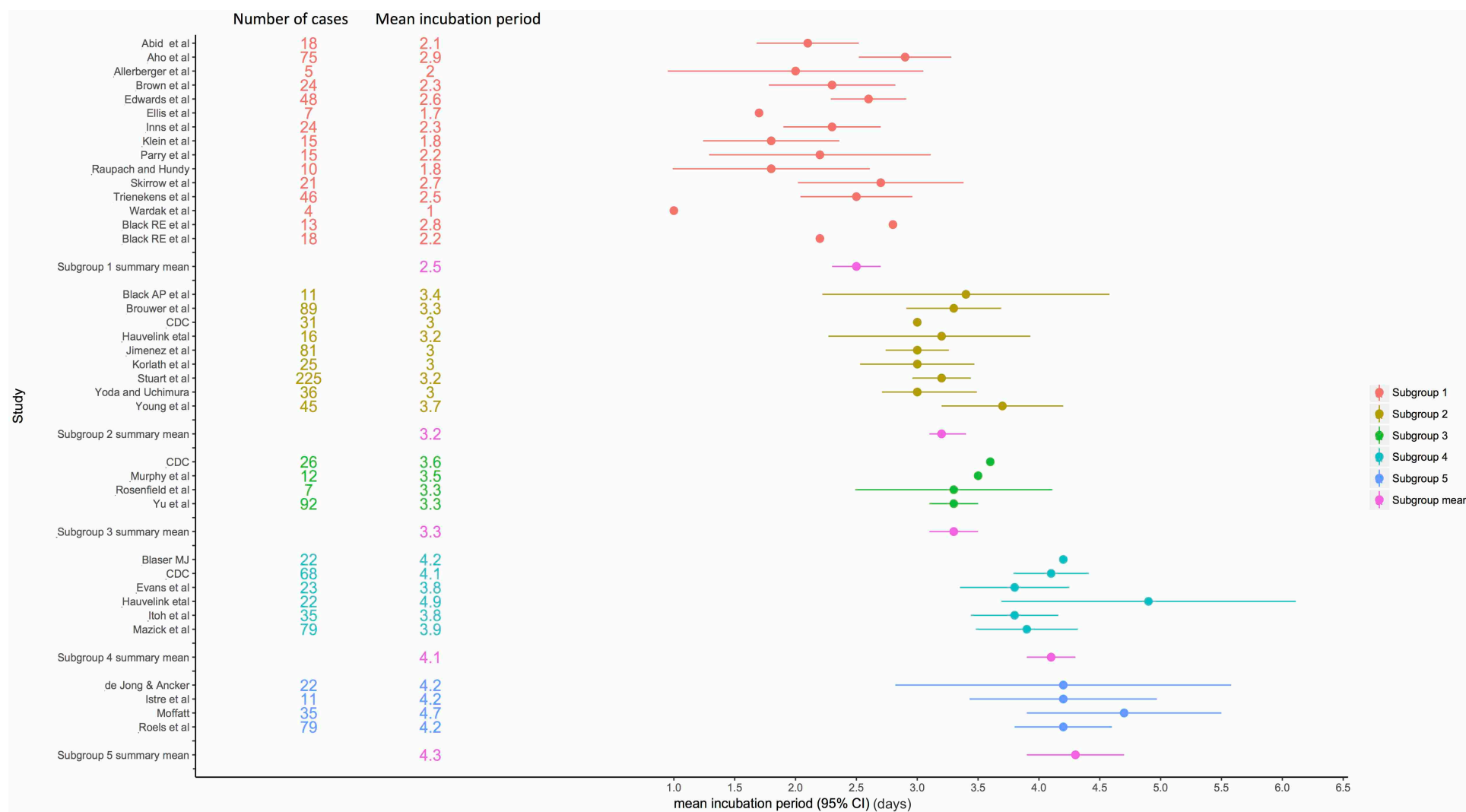


Figure 11 Forest plot showing mean incubation period and 95% CI in the review of the incubation period of Campylobacter

Table 9 Characteristics of studies within subgroups in the review of the incubation period of Campylobacter

Characteristics	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4	Subgroup 5
Age	63% adults	63% adults	50% adults	80% children	50% adults
Food vehicle	63.6% Poultry	50% poultry 25% dairy	100% poultry	60% dairy 20% poultry	50% poultry
Setting of outbreaks	55% Food service	25% farm 25% school 25% food service	50% food service 50% school	40% farm 20% school	50% food service 50% school
Severity of illness	63%	50%	50%	80%	100%
Duration of illness	0 – 24 days	0 – 20 days	1-6 days	0-18 days	1-9 days
Longest incubation period	10 days	8 days	5 days	11 days	14 days

4.3.7 Distribution of incubation period data

From the Cullen and Frey graph (Appendix 5), the distributions that were most suited to the data were gamma and beta. The beta distribution was further excluded because it requires the data to have an upper limit and a lower limit which may not be biologically plausible. Fitting the data to a gamma distribution by maximum likelihood, the resulting model parameters of scale and shape were 5.3 and 1.0 respectively. The goodness-of-fit plots (Figure 12) indicated that the data was well characterised by a gamma distribution.

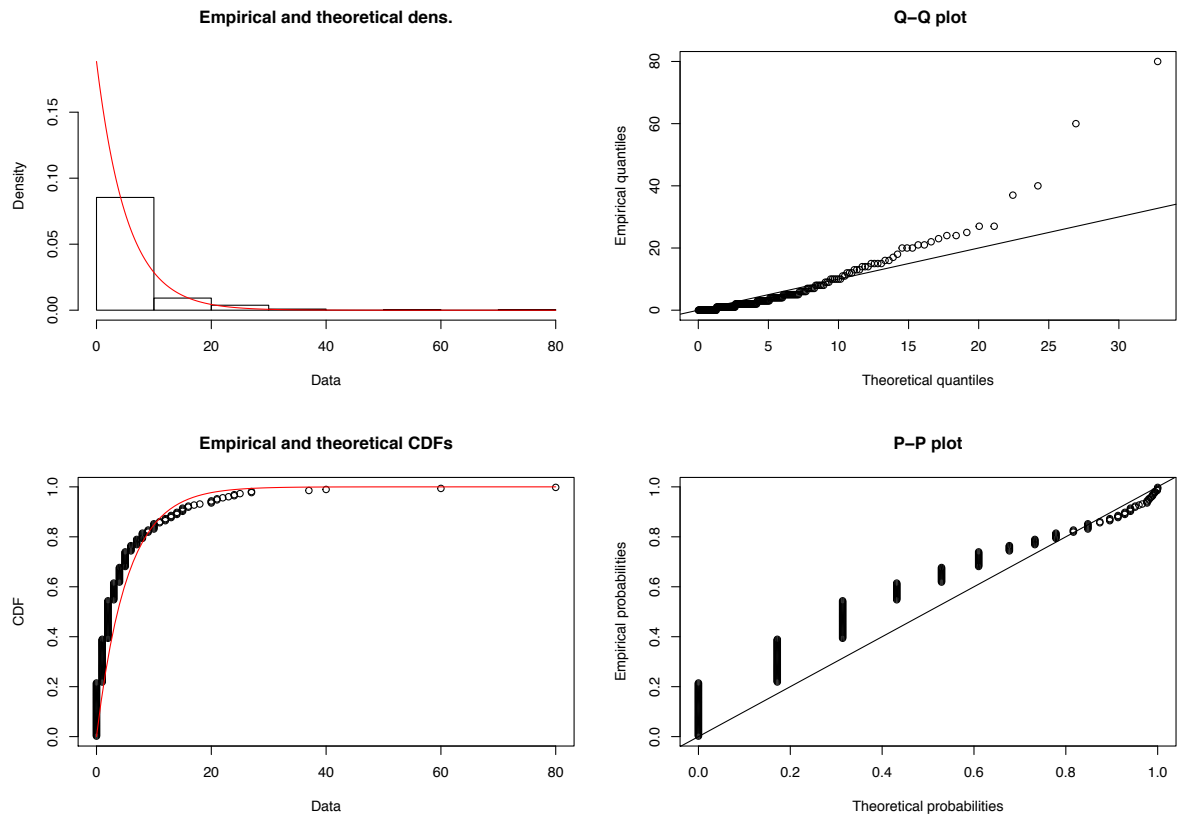


Figure 12 Goodness-of-fit plots for gamma distribution using maximum likelihood

4.4 DISCUSSION

Accurate estimations of the period between exposure and onset of illness for any infectious disease are essential to support evidence-based interventions in eliminating sources of infection. This review identified that the reported estimations of the incubation period of campylobacteriosis varied widely, even within subgroups of studies. The results of the I^2 and KS tests show that this variation is not due to chance, and there is an underlying pattern of variation. Visual inspection of Figure 9 and the results in Table 8 show that heterogeneity is not only in relation to mean incubation period, but also the shape of the distribution. From both regression analyses, age was identified as a factor that may influence the distribution of incubation period, with reported incubation period in outbreaks affecting children longer than those in mixed age groups. The age structure of cases of campylobacter has changed in recent years with older people becoming increasingly affected (128), and this population shift was reflected in this review where outbreaks investigated after year 2000 mostly involved adults or mixed age groups, while prior to year 2000, more outbreaks involving children were reported.

Furthermore, there appears to be some association between the subgroup characteristics and implicated food vehicle, setting of outbreaks and age of affected cases. However, these differences do not explain all of the variation in distributions of incubation period between subgroups. This may be due to other factors influencing distribution of incubation period that are not evident in the studies or inaccuracy of measurement and reporting. Host characteristics such as underlying medical conditions and immune response (129) as well as dose response have been known to affect infectivity and susceptibility to *Salmonella*, and may also influence the incubation period of other bacterial infections. These individual patient details have not been provided in the reports, so it is not possible to examine the effect of these factors.

The results of this review might not be generalizable to low and middle-income countries as majority of included outbreaks and experimental studies took place in high income countries in Europe and North America. Predisposing factors to campylobacteriosis in low and middle income countries, which might also influence incubation period, have been reported to be malnutrition and antimicrobial resistance (130). A further limitation of the current work is that case definitions varied between studies as authors used different criteria to define cases. The inclusion and exclusion of cases will therefore vary depending on the case definitions used, and this could also affect incubation period. However, all cases were identified at the onset of gastrointestinal symptoms including diarrhoea, vomiting and abdominal cramps, and all were in the context of a known outbreak or experimental study.

Outbreaks that mainly affected children were predominantly caused by consumption of raw milk or raw milk products and exposure was mostly during farm visits. This is similar to the report of Altekruuse et al (131). The incubation periods of outbreaks involving children were significantly longer than those of outbreaks involving adults or mixed age groups. A review of incubation period of infectious diseases in children reported a similar incubation period to my findings (67).

This study identified poultry and unpasteurised milk as the most common implicating food vehicles and are known causes of transmission(132,133). Studies have identified the presence of virulence genes in both poultry and dairy isolates (134). However, there is a disparity in the prevalence of *Campylobacter* in different food products (134) which may result in a variation in acquiring infection as well as incubation period. Also, some type of foods have been known to affect infectivity and thus potentially incubation period of pathogens by being either protective or enabling; an example is fatty food acting as a buffer to protect *Salmonella* from gastric acid (129).

Infectious dose may have a substantial effect on incubation period distribution, although this may not have varied substantially in the experimental studies included in this review. Studies

modelling the dose response of infectious diseases have reported a significant variation in the distribution of incubation period with dose (135,136). Human experimental studies of *Campylobacter* (58) and *Salmonella* (65) showed shorter incubation period where the challenge dose was higher. One of the reviewed studies reported a dose response relationship between the amount of milk consumed and onset of illness and severity, where cases drinking larger amounts of milk had shorter incubation periods and more severe symptoms (118). A dose response relationship was also reported in a non-foodborne outbreak involving an outdoor bike race where shorter incubation periods were seen in cases who reported ingesting larger quantities of mud (121). Another outbreak involving healthy military men who consumed at least four litres of untreated surface water during a military training exercise reported no dose response relationship between the quantity of water consumed and the severity of symptoms (137), however, there was no information on the relationship between ingested dose and incubation period. I was not able to analyse these relationships across the studies due to the lack of individual data related to dose and incubation time.

Host immunity could also influence the incubation period distribution as it determines if an exposure results in illness, and how long the process takes. The development of naturally acquired antibodies in response to a previous infection and the *C. jejuni* group antigen protects against subsequent illness (58), and may prolong incubation period if illness should occur.

It is worth noting that the bulk of the analyses has been carried out on a subset of studies included in the review from which raw data could be extracted. One problem I encountered in combining results of several studies was the different units of measurement used in reporting. Incubation periods were reported in hours, days or every two days. In order to combine the results, I converted all data to days, rounding up or rounding down where necessary. This could result in an over estimation where data was rounded up and an underestimation where data was rounded down and loss of precision for data from some

studies. Furthermore, using the online data extraction tool, Web Plot Digitizer, required manual selection of data points which is open to human error. Separating experimental studies and outbreak reports into relevant subgroups would have been an ideal way of analysing the data, however there was insufficient information to carry out these analyses, as there were four experimental studies and only two of these reported the mean incubation period.

Exclusion of non-English language articles is appropriate if processing these is inefficient as in this research team and is unlikely to produce bias. Bias would require that non-English papers are associated with different incubation period distributions in outbreaks. However, if there are few eligible studies, the translation and inclusion would be warranted. Furthermore, the study population is made up of cases that have been investigated as part of point source outbreaks where incubation period was not the main goal of investigation. This reduces the likelihood of publication bias and selection bias in the study population.

4.5 CONCLUSIONS

The results of this chapter confirm that incubation period in different outbreaks and experiments varied more than can be explained by chance, showed some clustering, and suggested that patient age may contribute to the variation. However, the information provided in the studies was not detailed enough to fully evaluate possible causes for these variations. The ideal data to support identification of factors affecting incubation period would be individual patient data across studies, including information such as underlying conditions, current medications and previous infections. In the absence of access to original individual patient data, reporting of outbreaks could allow better synthesis and meta-regression analysis. Although incubation period is not the main focus of outbreak reports they provide valuable natural experiments to describe incubation period distributions and

identify factors affecting this. Increased awareness of the value of this aspect of outbreak reporting can improve the presentation of data to support their use in evidence synthesis.

CHAPTER 5: INCUBATION PERIOD OF SHIGA-TOXIN PRODUCING *ESCHERICHIA COLI*: A SYSTEMATIC REVIEW AND META-ANALYSIS

ABSTRACT

Shiga toxin-producing *Escherichia coli* are pathogenic bacteria found in the gastrointestinal tract of humans. Severe infections could lead to life threatening complications especially in children under four years and the elderly. Understanding the distribution of the incubation period, which is currently inconsistent and ambiguous, can help in controlling the burden of disease.

This chapter outlines the process of undertaking a systematic review of outbreak investigation reports, extracting individual incubation data and summary estimates, testing for heterogeneity, classifying studies into subgroups with limited heterogeneity and conducting a meta-regression to identify factors that may contribute to the distribution of incubation period.

Twenty-nine studies were identified for inclusion in the review, and the resulting value of I^2 was 77% indicating high heterogeneity. Studies were classified into five subgroups with the mean incubation period ranging from 3.5 to 8.1 days. The length of the incubation period increased with age and decreased by 7.2 hours with every 10% increase in attack rate.

5.1 INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) is a pathogenic form of the *Escherichia coli* (*E. coli*) bacteria. It is common and benign in many organisms, but usually causes illness in humans (138). Symptoms generally include severe abdominal cramps, profuse diarrhoea which is often bloody, and vomiting (139). Symptoms could last between five to seven days (140), however severe infections can be life threatening and result in complications (139). A common complication of STEC infection is haemolytic uremic syndrome (HUS) which affects about five to ten per cent of ill persons (138,140).

There are several disease-causing STEC serogroups, but the most common is STEC O157 (141,142). Its reservoir is mainly cattle (143) and it is transmitted to humans through consumption of contaminated foods, faecal-oral transmission or cross contamination (142). Other STEC serogroups are collectively called non-O157 STEC. Numerous non-O157 strains have been identified however the most common serogroups associated with disease include O111, O26, O45, O103, and O145 (144,145).

Globally, STEC is estimated to cause nearly 3 million cases annually leading to approximately 200 deaths (26). Most reported cases of STEC O157 are sporadic (146), but due to the low infectious dose required for infection to occur (146), there is the potential for large outbreaks (147). Cases of non-O157 are relatively fewer and likewise, non-O157 related outbreaks are rare, however, large outbreaks have also been reported (148).

Incubation period is the time between exposure to the infecting pathogen and onset of clinical symptoms. Accurate knowledge of the distribution of incubation period is necessary for understanding the epidemiology of STEC, and is useful in outbreak investigations to distinguish between primary and secondary cases which is an important factor in its epidemiology as person to person transmission can occur (147). In an outbreak investigation

where the date of exposure is unknown, accurate knowledge of the incubation period distribution can help identify the potential source of contamination.

There are several conflicting reports on the incubation period distribution of STEC. The World Health Organization reports a range of three to eight days (138), the Centre for Disease Control and Prevention report three to four days after exposure (140) and cases in a few notable outbreaks have reported unexpectedly long incubation periods (57,149). Due to the ambiguity around the incubation period distribution of STEC, researchers have used other methods to identify the time of exposure in outbreak investigations (150).

In order to describe the incubation period distribution of STEC, examine the ambiguity around the reported incubation period and identify possible influencing factors, I undertook a systematic review of published and grey literature reporting the incubation period of STEC in exposed populations. I extracted reported individual patient incubation data and summary estimates and undertook a meta-regression of the extracted data. I tested for the presence of heterogeneity and attempted to explain observed variation by identifying influencing factors.

5.2 METHODS

5.2.1 Research questions and modified PICO elements

My research questions were:

1. What is the distribution of the incubation period of STEC infections in humans?
2. What factors affect the incubation period?

Population studies/Participants – Laboratory confirmed cases of STEC O157 and non-O157 in a point source exposure outbreak or continuous source outbreak where date of exposure and onset is known for each case or experimental study.

Probable cases of STEC infection based on clinical presentation and case definitions in the context of a point source exposure outbreak.

Infectious agent – Shiga-toxin producing *Escherichia coli* O157 and non-O157

Comparator – Host factors and any other factors such as ingested dose and food vehicle

Clinical characteristics such as HUS and bloody diarrhoea

Microbiological characteristics such as serotype and phage type

Outcome – Time from exposure to onset of clinical illness as described or defined by the reporting authors including diarrhoea, bloody diarrhoea, abdominal pain, vomiting, HUS

5.2.2 Search strategy

A systematic literature search for peer reviewed publications and grey literature reports of observational and experimental studies reporting incubation period was conducted in bibliographic databases (PubMed and Scopus) and Google Scholar. A customised grey literature search of the websites: www.cdc.gov, www.who.int, www.opengrey.eu and open Google search (screening only the first fifty items) was undertaken also using combinations of search terms (Table 10). All searches were carried out from 5 June 2017 to 3 July 2017. There was no restriction on the dates or language of articles returned. The reference lists of identified review papers were also screened to find other relevant studies where incubation period of STEC may have been reported.

Table 10 MeSH terms used in search strategy in the review of STEC

Search terms:

- I. outbreak* or experiment* or challeng* or case (tw)
- II. STEC or EHEC or VTEC or O157 or O104 or shiga-toxin (tw)
- III. dose-response or infectious-dose or threshold-dose or EC50 or CFU or MPN or incubation-period or incubation-interval or laten* or expos*
- IV. dose-response or infectious-dose or threshold-dose or incubation-period or incubation-interval or latent
- V. (epi-curve or epidemic-curve).mp.

Bibliographic database search: (I AND II AND III) OR (II AND V)

Google scholar search: (II AND IV) OR (II AND V)

Grey literature search: (II AND IV) OR (II AND V)

5.2.3 Selection process

Following the identification of likely relevant articles from the literature review, each article went through a selection process and quality assessment. These were done in the following steps:

- 1) Screening of titles and abstracts for articles reporting STEC infection
- 2) Screening of full text for reporting of incubation period
- 3) Review of full text to assess quality of incubation data reported and identify outbreaks with confirmed point source exposures or continuous source with individual exposure times.

The quality assessment process was done using a set of criteria (Table 1)(151)(141), to evaluate the quality of incubation period data reported and identify outbreaks or cases with definite exposure times. The selection and assessment process were done by two reviewers independently assessing the studies. Discrepancies were discussed until reviewers arrived at a consensus.

5.2.4 Data extraction

Data were extracted from all included studies according to a pre-determined proforma (Table 11). When available, individual incubation period data were extracted from text and tables. Where an epidemic curve was provided, the raw incubation period data were extracted using the online tool, Web Plot Digitizer (96). Summary data, such as the mean, median and mode, were also extracted or calculated from the raw data where necessary. Raw data were used in the test for heterogeneity and the subgroup analysis. The summary data were used in the meta-regression analysis as explanatory variables were only available for outbreaks and not individuals. The unit of incubation data reported and extracted was days. For point source outbreaks, or outbreaks with continuous exposures where the total number of people exposed was observed and reported, the attack rate was determined.

Table 11 Details of information extracted in the review of STEC

Section	Information to be collected
General information	- Year of publication - Title of article - Authors - Type of publication (journals, conference abstract, grey literature, etc.) - PubMed ID (where applicable)
Study characteristics	- Year of study - Study design (cohort, case-control, experimental, case series) - Country of study - Age distribution - Comments on method or quality of study
Pathogen characteristics	- Infectious agent - Serotype - Subtype - Toxin type - Phage type
Outcome data/ results	- Case definition - Reported incubation period (individual data, mean, median mode and range) - Derived or calculated summary estimates incubation period (raw data extracted) - Source of calculated data (epidemic curve or author description)

Other outcome data	- Time between onset of symptoms and bloody diarrhoea or HUS - Duration of illness
Factors that could affect incubation period	- No of exposed cases - No of people affected - Setting - Mode of transmission - Food vehicle (for foodborne infections only) - Patient characteristics (e.g. previous infection or treatment, underlying illness, age)
Any other relevant information	- Any other relevant information

5.2.5 Descriptive analyses

The extracted data were summarised according to the reported characteristics. Frequencies and percentages were calculated for year of study, study design, geographical region of study, serotype, toxin type, mode of transmission and food vehicle, where applicable.

The extracted individual incubation period data were used to plot histograms of the incubation periods and re-create the epidemic curves of the outbreaks. The epidemic curves were plotted on a uniform x-axis representing the incubation period from one to twenty days and an individual y-axis indicating the number of cases in each outbreak.

5.2.6 Statistical analyses

I tested for the presence of heterogeneity using the individual incubation data, and the pattern of heterogeneity was investigated. Factors that may explain heterogeneity were also identified using the summary statistics and available outbreak characteristics. Statistical

analyses were done using the statistical software R version 3.2.3 (2015-12-10) – “Wooden Christmas Tree” (127).

5.2.6.1 Testing for heterogeneity

The Friedman statistic was derived and the value of I^2 calculated. I also performed a Kolmogorov Smirnov test (KS test) to compare the cumulative distributions between studies. A bootstrapped version of the function was applied estimating the sampling distribution by re-sampling the population 10,000 times. I calculated the proportion of p-values less than 0.05 and the probability of obtaining the observed proportion.

5.2.6.2 Identifying factors that explain heterogeneity

In order to examine the relationship between the outbreak characteristics and the mean incubation period, a linear regression analysis was performed. A generalised linear model, with a gamma family function and a link identity, was fitted to account for the skewness of data. The mean incubation period was the dependent variable and the outbreak or host characteristics were the explanatory variables. The effect of the explanatory variables on the mean incubation period was examined first using a univariate model, and then by building a multivariable model with variables that had a significant association (p-value <0.05) to test for confounding.

5.2.6.3 Identifying subgroups of studies/Describing pattern of heterogeneity

Following the confirmation of heterogeneity, the data according to subgroups was explored. Using the p-value from the KS test, a hierarchical cluster analysis was conducted to produce a graphical representation of the dissimilarity matrix between the studies. The output was a dendrogram showing a compact visualisation of the dissimilarity matrix. Studies on the dendrogram were classified into subgroups by applying pragmatic adjustments to the

significance level and subtracting the adjusted p-value from 1 ($1 - \alpha$) to derive the cut-off point for identifying subgroups of studies with limited evidence of heterogeneity.

5.2.7 Subgroup analyses

The individual incubation data of studies in a subgroup was collated and summary statistics and outcome measures were derived for each subgroup. A forest plot was created to show the distribution of the mean incubation period and the corresponding 95% confidence interval for studies with individual incubation data.

5.2.8 Risk of bias

The data were tested for 'small-study effect' by creating a funnel plot which graphically represented the relationship between sample size and incubation period.

5.2.9 Determining the distribution of incubation period data

I aggregated the extracted raw incubation data and created a Cullen and Frey plot in order to identify distributions that best fitted the data and exclude other distributions.

5.3 RESULTS

5.3.1 Search strategy and selection process

A total of 1980 unique articles were retrieved from both bibliographic databases, 840 from Google Scholar and 1279 articles were identified from the grey literature search. After screening for relevance and removing duplicates identified from all three sources, 2059 articles were excluded (Figure 13). An additional 26 articles were identified from searching through reference list of relevant review papers resulting in 2066 articles for further screening. Excluding articles that did not report incubation period resulted in 42 articles for full text review and eligibility screening. A further fourteen articles were excluded (Appendix 6) as they did not meet the eligibility criteria resulting in 28 articles for inclusion and data extraction, one of which reported two outbreaks. All of the included studies were outbreak reports, and 22 of these reported individual incubation period data available for extraction (Appendix 7). Some outbreaks were part of larger outbreaks. Three outbreaks (152–154) within the German O104 outbreak, for which incubation time could be extracted, were included. It was not possible to include all cases in the 2011 German outbreak as incubation time was not known due to uncertain dates of exposure.

5.3.2 Descriptive analyses

Of the 29 studies included in our review, 75.9% (22/29) were of serotype O157. All but one of the studies reported the onset of either bloody diarrhoea or HUS. Nine outbreaks (31%) each reported bloody diarrhoea alone or HUS alone and ten outbreaks (34.5%) reported both HUS and bloody diarrhoea (Table 12).

Outbreaks involving mostly children accounted for about half (51.7%; 15/29) of the included studies and settings reported comprised of farm visits (5/15), schools (4/15), and outdoor settings (4/15) such as swimming pools, camping and exposure to surface water. Outbreaks

involving mostly adults reported settings such as private parties (6/13) and nursing homes (4/13).

Foodborne transmission accounted for 69% (20/29) of the outbreaks of which 40% (8/20) were private parties, and 20% (4/20) took place in either a nursing home or a school. Non-foodborne transmission accounted for 31% (9/29) of outbreaks, of which 44% (4/9) were associated with farm visits. The most commonly identified food categories in foodborne outbreaks were vegetables (35%; 7/20) and red meat (20%; 4/20) (Table 12).

The funnel plot showed no evidence of small-study effect. It was a symmetrical funnel with small sample studies reporting both short and long incubation periods (Figure 14). The re-created epidemic curves from the extracted individual data showed variation in the distribution of the incubation period (Figure 15).

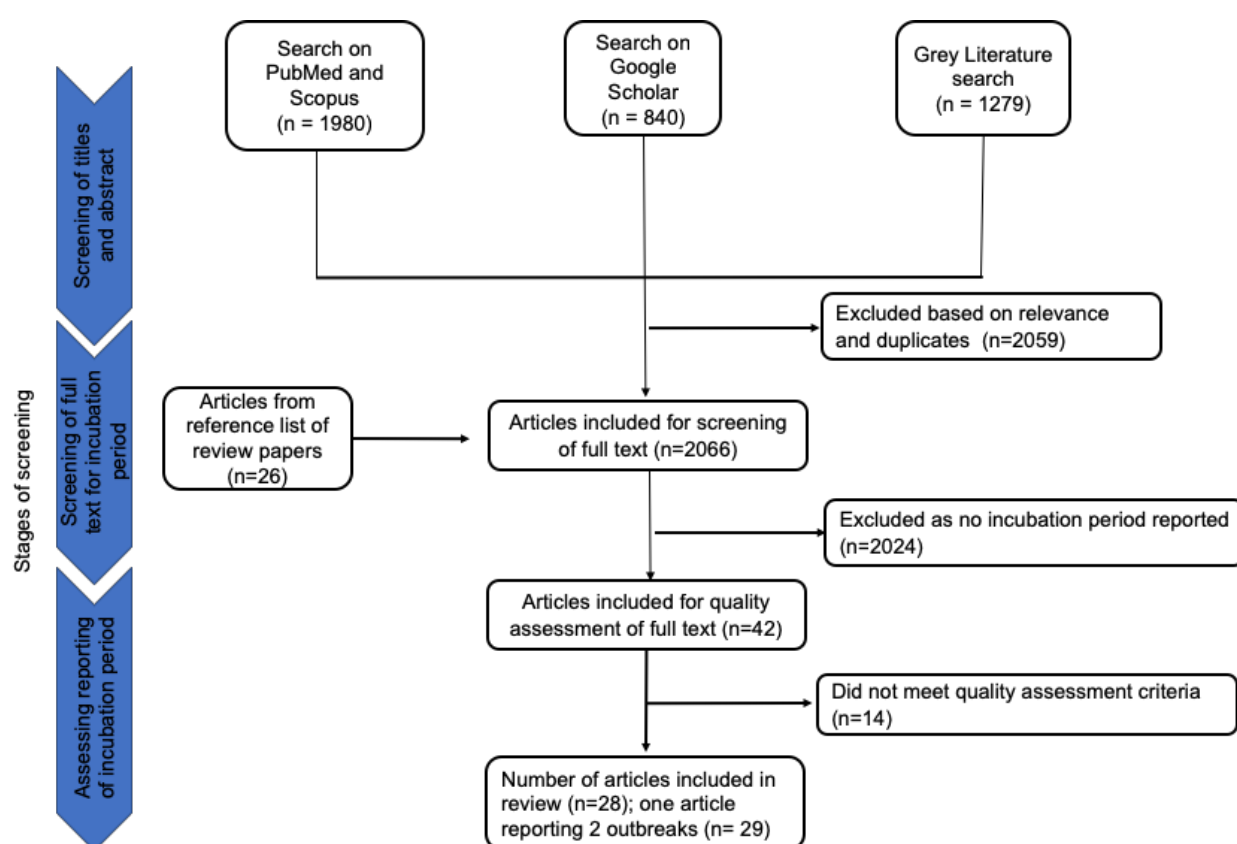


Figure 13 Flowchart of study selection process in the review of STEC

Table 12 Characteristics of included outbreaks in the review of STEC

	Number	Proportion
Total number of outbreaks	29	
Year of study		
Before year 2000	16	55.2
2000 and later	13	44.8
Study design		
Case control study	9	31.0
Case study	1	3.4
Cohort study	14	48.3
Descriptive	5	17.2
Age group		
Children	15	51.7
Adults	13	44.8
Mixed	0	0.0
Unknown	1	3.4
Region of study		
Europe	16	55.2
North America	8	27.6
Asia	5	17.2
Serotype		
O157	22	75.9
O104	4	13.8
O127	1	3.4
O145 & O26	1	3.4
O103	1	3.4
Toxin type		
VT1 & VT2	9	31.0
VT2 alone	8	27.6

VT1 alone	3	10.3
Unknown	9	31.0
Clinical outcome		
Bloody diarrhoea	9	31.0
HUS alone	9	31.0
Both	10	34.5
Unknown	1	3.4
Foodborne transmission	20	69.0
Private party	8	40.0
Nursing homes	4	20.0
School	4	20.0
Community	1	5.0
Farm visit	1	5.0
Restaurant	2	10.0
Non-foodborne transmission	9	31.0
Farm visit	4	44.4
Laboratory acquired	1	11.1
Outdoor activity	2	22.2
Surface water	1	11.1
Swimming pool	1	11.1
Food vehicle		
Vegetables	7	35.0
Red meat	4	20.0
Dairy	3	15.0
Poultry	2	10.0
Others	2	10.0
Unknown	2	10.0

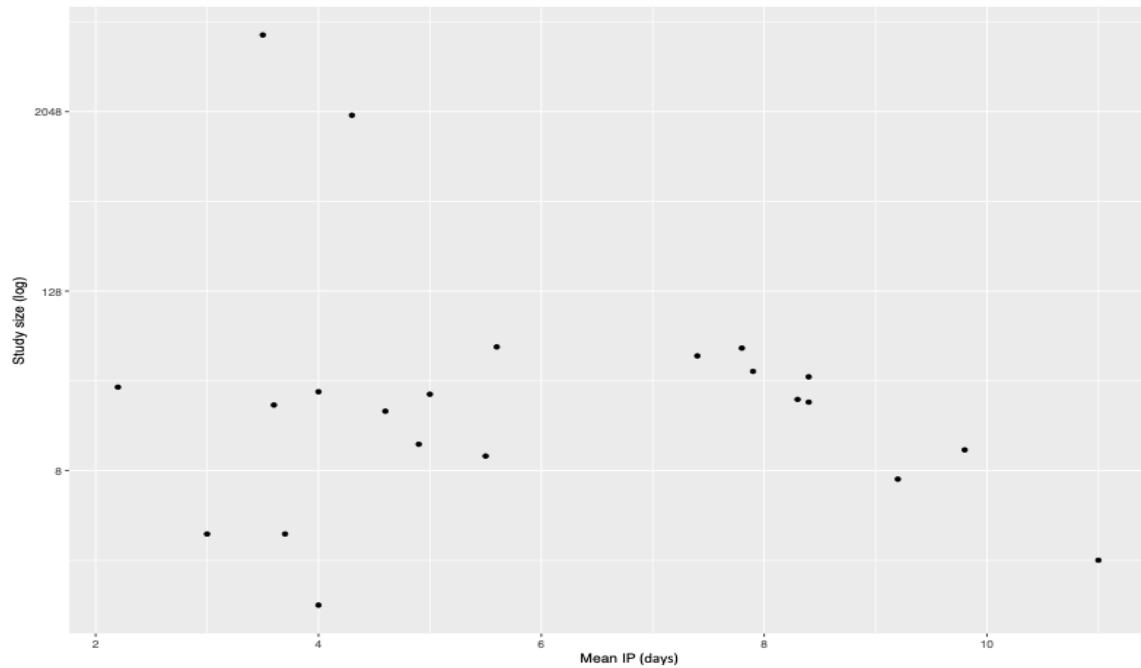


Figure 14 Funnel plot showing relationship between sample size and mean incubation period in the review of STEC

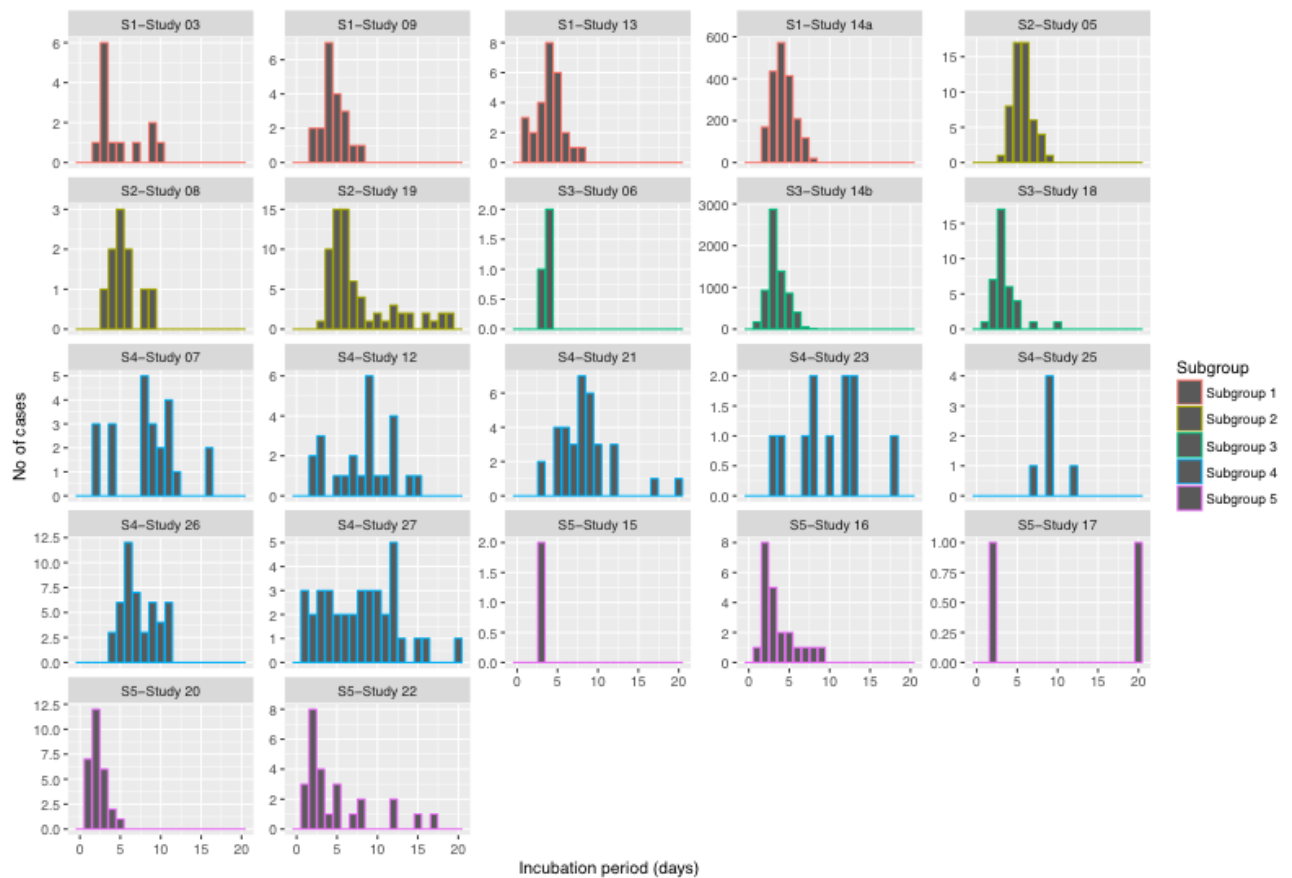


Figure 15 Collated epidemic curves of included studies in the review of STEC

5.3.3 Test for heterogeneity

The results of the Friedman statistics (<0.001) and I^2 (77%) indicated high heterogeneity between the studies. From the KS test, 56% (143/253) of the resulting p-values were less than 0.05 and the probability of obtaining this proportion by chance was <0.0001 .

5.3.4 Factors that may explain heterogeneity

The results of the regression analysis showed outbreaks involving mostly children and attack rate as factors that may influence the mean incubation period. From the univariate analysis, outbreaks with children reported an incubation period of 2.7 days shorter than outbreaks involving adults (p-value – 0.01). A 10% rise in the attack rate resulted in a reduction in incubation period by 0.3 days (p-value – 0.03). The results were similar in the multivariable analysis; however, the p-value indicated a borderline significant association with the mean incubation period (p-value- 0.06 for both) (Table 13).

Table 13 Results of generalised linear regression model in the review of STEC

Variables	Univariate model		Multivariable model	
	Difference in IP (95% CI)	p-value	Difference in IP (95% CI)	p-value
Age				
Adult	Reference			
Children	-2.7 (-5.1, -0.8)	0.01	-2.1 (-4.3, -0.02)	0.06
Attack rate	-0.03 (-0.1, 0.0)	0.03	-0.03 (-0.06, 0.01)	0.06
Setting				
Other	Reference			
Farm	1.4 (-4.6, 4.9)	0.4		

Nursing home	3.2 (-3.0, 8.3)	0.1		
Outdoor	1.1 (-5.1, 5.8)	0.6		
Private party	3.1 (-2.8, 6.3)	0.1		
School	-0.6 (-4.6, 6.7)	0.7		
Restaurant	1.5 (-6.5, 2.1)	0.4		
Mode of transmission				
Foodborne	Reference			
Non-foodborne	-0.6 (-2.7, 1.9)	0.5		
Serotype				
Non O157	Reference			
O157	-1.19 (-3.9, 0.9)	0.3		
Toxin type				
VT1	Reference			
VT2	1.5 (-2.2, 7.5)	0.3		
VT1 & VT2	1.5 (-2.1, 4.7)	0.3		
Bloody diarrhoea				
No	Reference			
Yes	-0.14 (-2.7, 1.9)	0.9		
HUS				
No	Reference			
Yes	0.3 (-1.9, 2.3)	0.7		

5.3.5 Identifying subgroups of studies

The output of the clustering analysis produces a dendrogram visualising the dissimilarity matrix. As a result of the pragmatic adjustment made to the significance level, the corrected p-value was 0.002, and subtracting this from 1 resulted in a cut-off point of 0.998. Applying this cut-off point to the dendrogram resulted in five subgroups of studies with limited evidence of heterogeneity between them. These consisted of a subgroup of seven studies, a subgroup with five studies, a subgroup of four studies and two subgroups with three studies each (Figure 16).

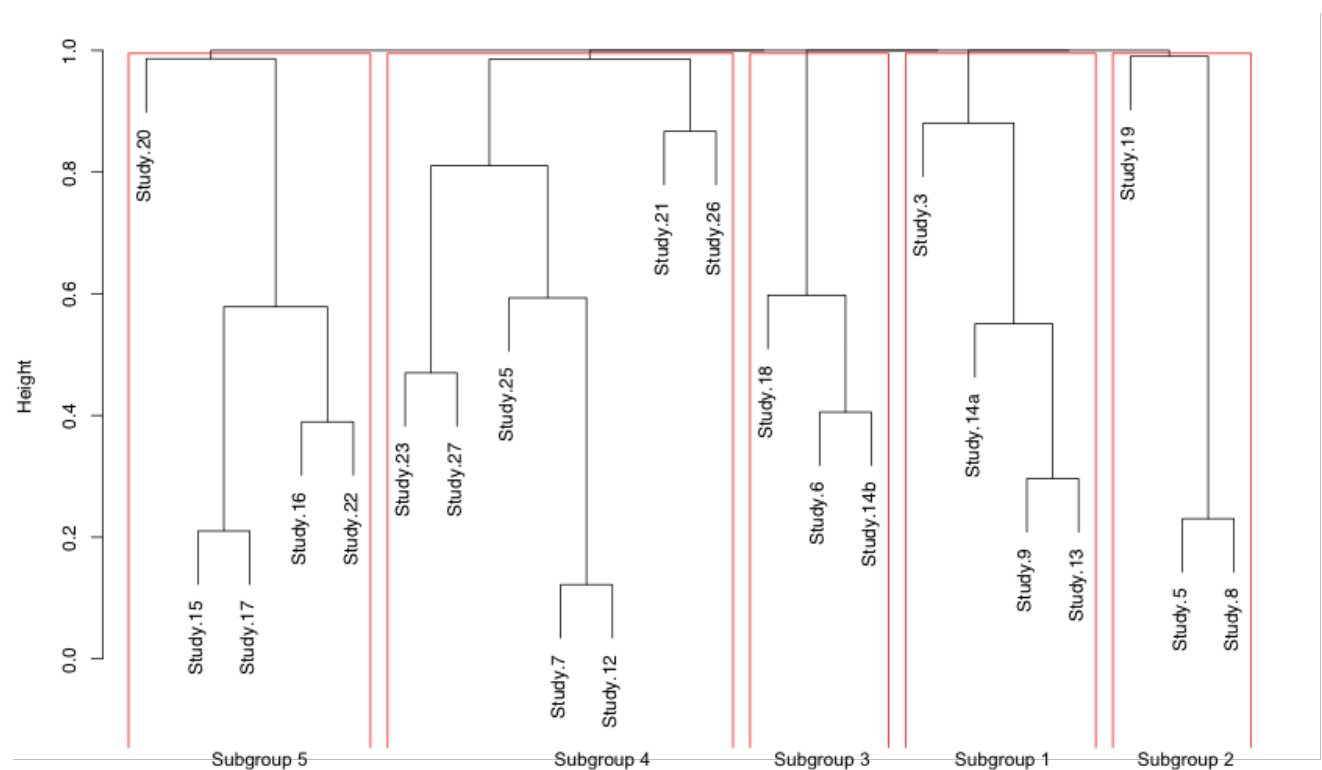


Figure 16 Dendrogram showing dissimilarity matrix and associated subgroups in the review of STEC

5.3.6 Summary of subgroup analysis

The mean incubation period between subgroups varied from 3.5 to 8.1 days, and differed significantly between subgroups 1 to 4 (Figure 17). Subgroups 2 and 4 had significantly

longer mean incubation periods of 6.7 days and 8.1 days respectively compared with the other subgroups (Table 14). The variance, skew and kurtosis also differed, and were larger for subgroups with smaller sample sizes. Some variation was observed within subgroups particularly in outbreaks reporting symptoms of bloody diarrhoea and HUS and the outbreak setting.

Two out of three outbreaks in subgroup 2 involved mostly adults and all outbreaks in subgroup 4 involved mostly adults. The distribution of serotypes in the subgroups was diverse, although all outbreaks involving serotype O104 were clustered within subgroup 4. Severity of symptoms was similar across subgroups as outbreaks in all subgroups reported both bloody diarrhoea and HUS (Table 14).

5.3.7 Distribution of incubation period data

The Cullen and Frey graph (Appendix 8) indicated that the data were most suited to a beta distribution. However, this distribution is restricted to lower and upper limits respectively which may not be biologically plausible. Further investigation of the use of such families of distributions will be necessary, but I limited this study to the gamma distribution which is more usually used to model incubation periods.

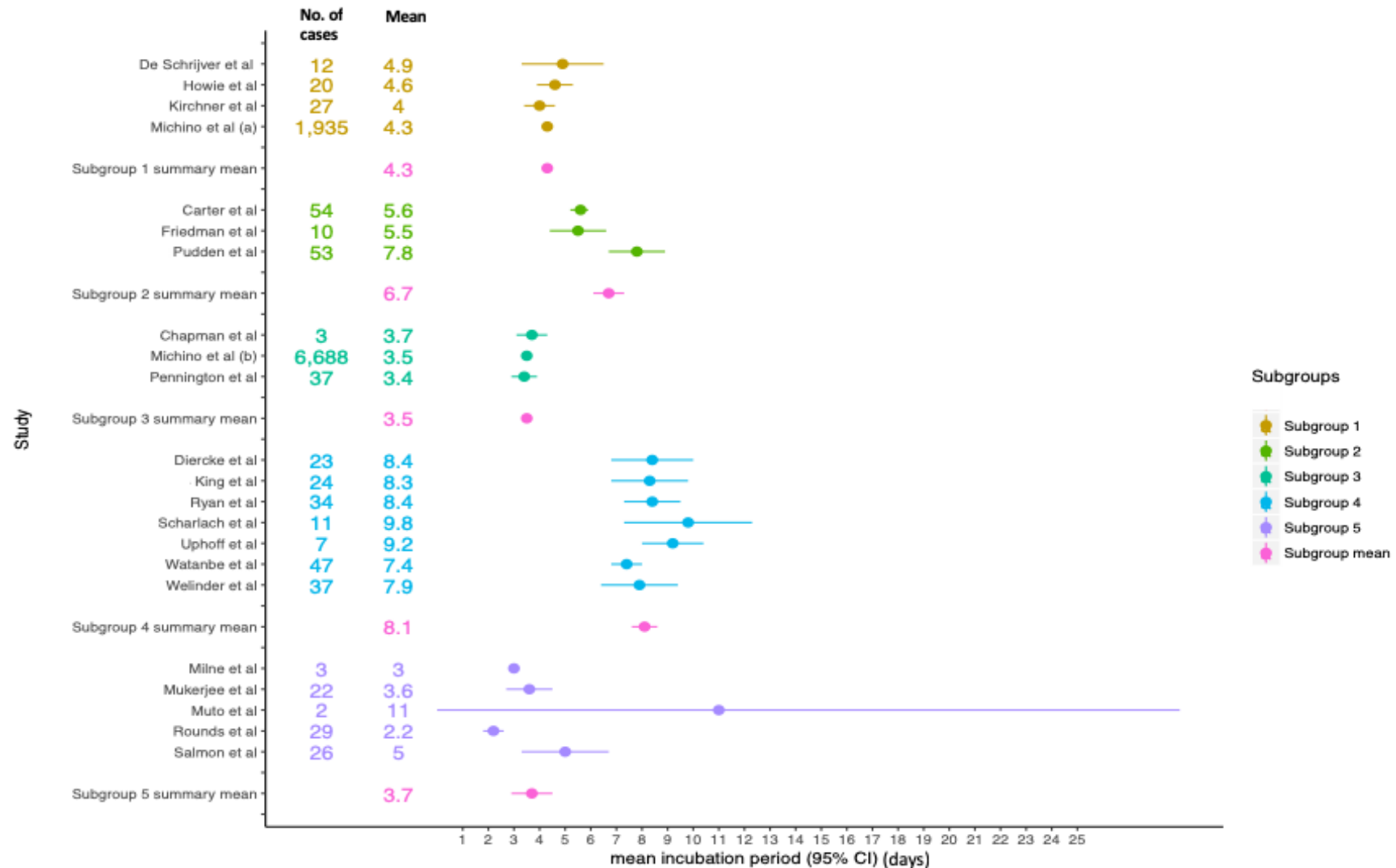


Figure 17 Forest plot showing mean incubation period and subgroup summary mean in the review of STEC

Table 14 Summary and characteristics of studies within subgroups in the review of STEC

	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4	Subgroup 5
Number of studies	4	3	3	7	5
Sum of cases	1995	133	6728	182	80
Mean Incubation period (95% CI)	4.3 (4.2 - 4.4)	6.7 (6.1 - 7.3)	3.5 (3.4 - 3.5)	8.1 (7.6 - 8.6)	3.7 (2.9 - 4.5)
Median Incubation period	4	6	3	8	2.5
Variance	1.9	11.2	1.4	12.8	12.7
Skew	0.4	2.1	0.6	0.5	2.7
Kurtosis	-0.1	4.3	0.3	0.7	7.9
Age	100% children	67% adults	67% children	100% adults	100% children
Bloody diarrhoea	50%	67%	67%	71%	80%
HUS	75%	67%	33%	71%	60%
Setting of outbreaks	25% Private party, 25% School, 25% Farm	67% Nursing home	33% Private party, 33% School, 33% Farm	71% Private party	40% Farm, 40% School
Longest incubation period	10 days	19 days	10 days	20 days	20 days

5.4 DISCUSSION

I have undertaken a systematic review of published and grey literature and identified articles reporting precisely estimated data on the incubation period of STEC. I extracted the reported data and due to the presence of heterogeneity, classified studies into five subgroups for analysis among which the mean incubation time varied as well as other measures of distribution such as the median and variance. The mean incubation period of the subgroups ranged from 3.5 to 8.1 days. Age and attack rate were identified as factors that influence incubation period.

In this review, the length of incubation period increased with age as outbreaks involving mostly children reported shorter incubation periods. This contrasts with the findings of Werber et al (155) where incubation period decreased with age in a single O104 outbreak using individual patient data. Children are more at risk of STEC infections (26) for a number of reasons, some of which could also influence the incubation period. However, this study could only perform ecological analysis due to the lack of individual patient data and additionally lacked information to assess many possible explanatory factors that may explain the association of age with the distribution of incubation period.

The attack rate of a disease and incubation period have been reported to have an inversely proportional relationship such that a higher attack rate results in a shorter incubation period in a study of *Salmonella* outbreaks (156). This study showed that for every ten percent increase in attack rate, the incubation period was shortened by 7.2 hours or 0.3 days. Factors such as virulence of the pathogen, the infectious dose and host susceptibility which alter the attack rate of a disease in a population (157–160) and also the incubation period of the disease (161) may have contributed to this association.

I did not find statistical evidence for association of other features with incubation time: setting, mode of transmission, symptoms, toxin type and serotype. STEC O157 is the most

commonly reported serogroup (141) and was also the dominant STEC described in this review accounting for 22 outbreaks out of 29. The O104 serogroup that caused the German outbreak accounted for 3 outbreaks, hence, the paucity of outbreaks caused by non-O157 serogroup may be responsible for the non-significant association between serogroup and incubation period. Despite this, all outbreaks associated with the O104 serogroup were clustered in one subgroup, which had the longest incubation period. Mostly reporting serotype and toxin type, the outbreaks studied lacked more detailed microbial characterisation. As routine STEC genome sequencing increases, testing of a wider range of pathogen factors for association with incubation period will be possible. The observed clustering of the O104 outbreaks within the same incubation time subgroup (Figure 16) (152–154,162), where outbreaks shared both a pathogen lineage and a transmission route through contaminated beansprouts supports the potential importance of these factors in affecting incubation period.

The variability of incubation time distribution between outbreaks, and the large number of outbreaks with incubation times longer than those cited in reference materials (138,140) is striking. The mean incubation time across the seven outbreaks in subgroup 4 was longer than the standard ranges proposed for individual cases (138,140). Even acknowledging that three of these seven were from those outbreaks (152–154) within the large German O104 outbreak (57) with identifiable incubation times and one from a related O104 outbreak in France (162), this large cluster of outbreaks with longer than expected incubation times highlights the risks associated with restricting investigation to exposures occurring within expected incubation time distributions. A consequence is the need for outbreak investigators to consider a wide range of potential incubation times. This chapter offers an evidence base of the scale of variation to inform future editions of reference materials.

Although I was able to summarise data across the published literature, this study is based on very limited data compared to the large burden of disease of STEC and the numerous outbreaks investigated and reported (142,147). The majority of STEC cases are either

sporadic (163), and some are part of continuous source outbreaks (164–167) where it is difficult to identify exposure time and therefore difficult to calculate incubation periods accurately. Many studies were excluded because, although they appear to have been gathered, these data were not reported in a way that allowed suitable data extraction. Standard approaches to reporting incubation period data in outbreak reports should be developed and would be useful for better understanding incubation periods. The same is no doubt true for other questions, including the effectiveness of control measures, where the natural experiments offered by outbreaks could contribute to evidence based practice if well reported and collated.

One reporting feature of importance to incubation time is the case definition used. Where data for both were available we observed a period of four to seven days between onset of diarrhoea and onset of HUS, similar to that reported by Razzaqq et al (168). In this review, most of the case definitions used by the authors included diarrhoea or bloody diarrhoea without HUS, while a few included all three clinical outcomes. I observed that studies reporting all clinical outcomes including HUS reported longer mean incubation period than other studies, however, I did not identify a significant association between HUS and the length of incubation period as also reported by Werber et al (155). I suspect that some of the heterogeneity observed in the incubation period distribution across outbreaks may be explained by the differences in the case definitions. Reporting with more specific case definitions and even separation of results across cases based on different symptoms is thus one example of a feature of standard reporting that could promote better evidence synthesis.

Even with detailed and standardised reporting of outbreaks there are limitations to using the published literature to study factors associated with incubation time. For these questions, individual patient data allowing analysis of host factors such as pre-morbidity, ongoing medications and dose is required alongside outbreak level characteristics such as mode of transmission and pathogen characteristics. This raises questions on how best to capture, store, and make accessible data from individual outbreaks in a way that will allow joint

analysis and align with the increasing trend toward individual patient data meta-analysis in the synthesis of randomised controlled trial evidence.

The studies included in this review were predominantly outbreak reports where incubation period was not the main goal of investigation and the population being studied were cases being investigated as part of the outbreak. Therefore, there is very little possibility of encountering publication bias or selection bias due to the reported incubation period. Furthermore, I found no evidence of small-study effect. Although protecting against publication bias, this lack of focus on sharing information on incubation period no doubt contributed to the exclusion of studies, with the majority of studies identified not meeting the quality assessment criteria for reporting of incubation period. Extracting individual data involved the use of an online data extraction tool and a manual process of selecting each data point which is open to human error and could alter the reported incubation period, but these errors are likely to be small. Results are therefore not likely to be substantially biased, although based on a restricted number of studies reporting usable data.

5.5 CONCLUSIONS

In conclusion, this chapter confirmed that the incubation period varied across outbreaks identifying several subgroups some of which had incubation periods far longer than expected based on standard estimates. I identified age and attack rate as factors that may influence the distribution of incubation period. However, there was insufficient information to explain the relationship between these factors and incubation period or to study the impact of many individual patient level factors on incubation period. This chapter highlights both the opportunities afforded for information synthesis across outbreaks to support evidence-based practice and the challenges to be overcome to optimise reporting and ultimately support combined analysis across outbreaks if the full potential of these natural experiments is to be exploited.

CHAPTER 6: ANALYSIS OF INDIVIDUAL PATIENT DATA TO UNDERSTAND THE INCUBATION PERIOD DISTRIBUTION OF SHIGA-TOXIN PRODUCING *ESCHERICHIA COLI*

ABSTRACT

Shiga-toxin producing *Escherichia coli* (STEC) is a pathogen that can cause bloody diarrhoea and severe complications. Outbreaks are commonly reported with several having unusually long incubation period. Understanding the incubation period distribution as well as influencing factors will help in identifying exposure times during outbreak investigations.

I extracted individual patient data of STEC cases associated with outbreaks with a known source of exposure in England and Wales between 2009 and 2016. Data was retrieved from the enhanced STEC surveillance system. The incubation period was derived and cases were described according to the patient and outbreak characteristics. I tested for heterogeneity between the outbreaks using Kolmogorov-Smirnov test and described the pattern of heterogeneity using hierarchical cluster analysis. A multi-level regression model was employed to examine the relationship between patient characteristics such as age, gender and reported symptoms; and outbreak characteristics such as mode of transmission with the incubation period.

A total of 205 cases from 41 outbreaks were included in the study. The median incubation period was four days. There was evidence of heterogeneity between outbreaks, and from the clustering analysis, the clustering of outbreaks into subgroups was partly driven by the mode of transmission. Cases reporting bloody diarrhoea reported shorter incubation periods compared with cases without bloody diarrhoea.

6.1 INTRODUCTION

Shiga-toxin producing *Escherichia coli* (STEC) infection is a zoonotic pathogen that causes gastrointestinal illness. The main routes of transmission are foodborne, direct or indirect animal contact, and person to person spread (142). Common symptoms include diarrhoea, nausea and vomiting (139) and some cases report symptoms of haemorrhagic colitis, including bloody diarrhoea and abdominal pain (25). Between 6% to 14% of cases develop haemolytic uremic syndrome (HUS), a serious complication of STEC infection (169).

Outbreaks of STEC are frequently reported in England and Wales, with an average of 13 outbreaks reported annually between 2009 and 2016 (147). Investigating outbreaks, which contributes to reducing the burden of STEC infections, usually involves collecting information on risk factors through patient interviews and questionnaires, either as part of routine surveillance or through epidemiological studies and/or environmental investigations. The information is collected for a set time period prior to clinical illness, aiming to cover the period a case may have been exposed to the pathogen. Conducting these interviews, informed by a good understanding of when exposures leading to disease are likely to have occurred, supports effective and efficient investigation.

Incubation period is the time between exposure to the infecting pathogen and the onset of clinical illness. Accurate knowledge of this parameter is useful in narrowing down the possible time of exposure, excluding secondary cases and also declaring the end of an outbreak (48). It is also useful in understanding the pathogenesis of STEC and estimating the possible extent of the spread of infections. In the large notable STEC O104 outbreak in Germany (57), the incubation period was longer than expected resulting in challenging epidemiological investigations particularly when collating relevant exposure details.

Despite its importance, available reports on the incubation period of STEC are conflicting, such as the differing expected distributions proposed by the World Health Organization and

the Centre for Disease Control and Prevention of three to eight days and three to four days respectively (138,140). Additionally some outbreaks have reported incubation periods of a median of nine days (162). Incorrect estimations of the incubation period may result in formulating inaccurate case definitions and wrongly defining exposure times.

Systematic reviews of outbreak reports and experimental studies have proved invaluable in describing the incubation period distribution of gastrointestinal pathogens, and identifying possible influencing factors (151,159). However, some patient factors, such as patient demographics and reported symptoms, could not be examined as they were not available in published outbreak reports at the individual level. Analysing individual patient data (IPD) provides an avenue to study these additional factors. IPD can be extracted from both sporadic cases and cases associated with outbreaks. The source of infection is more likely to be reliably identifiable in outbreaks than for sporadic cases, making these data a useful resource to study incubation period. Unusually, England has a national system to collate individual case data for STEC cases and outbreaks allowing analyses of these data.

In this chapter, I extracted individual patient data from outbreaks with a reported source of infection reported in England and Wales between 2009 and 2016. I calculated the incubation period for each case and analysed both patient and outbreak data to identify the distribution of incubation period and factors that may influence the incubation period of STEC.

6.2 METHODS

6.2.1 Case ascertainment

Routine enhanced surveillance of STEC has been ongoing in England and Wales since 1 January 2009 (141). The local laboratories are expected to report presumptive isolates to Public Health England Centres (PHEC). As part of a routine follow-up, the PHEC administers a standardised enhanced surveillance questionnaire (ESQ) (170) (Appendix 9) to each patient. Clinical and epidemiological data collected by the ESQ include: demographic information; relevant occupation including food handlers, health care workers and child carers; date of illness onset; clinical condition including HUS progression; details of potential exposures including travel, food and water consumption, contact with animals and environmental factors; and outbreak status. In addition, cases are required to indicate the symptoms they experienced from a standard set of symptoms. Completed questionnaires are then forwarded to the gastrointestinal department of Public Health England (PHE) and entered into the National Enhanced STEC Surveillance System (NESSS). Cases associated with outbreaks where the source was identified were included in the study. Additional outbreak information such as outbreak reports and outputs of investigations were also sought and cases were included where there was information at the outbreak or individual level allowing identification of their exposure date.

6.2.2 Comparing characteristics of excluded and included cases

The cases excluded from the study were described according to available characteristics including age, gender, ethnicity, geographical region of outbreak, reported symptoms and mode of transmission. A chi-square test for equality of proportions was calculated to compare the proportion of cases in the excluded group to the proportion of cases in the included group. The significance level was set at a p-value of <0.05.

6.2.3 Microbiological methods

Confirmation and serotyping of STEC was done at the PHE Gastrointestinal Bacteria Reference Unit (GBRU). Strains of serogroup O157 were further differentiated by phage type (PT) (141).

6.2.4 Calculation of incubation period

Date of illness onset, for the purpose of this study, was defined as the first day of symptoms as reported by the patient, which is routinely collected in the ESQs.

The date of exposure was defined as the day the patient reported to have come into contact with the identified source of infection, such as a contaminated food item or infected animal.

The date of exposure is not asked as a routine question on the ESQ, but rather, the question is asked as to whether exposure occurred during the seven days before illness. Where the exposure date was unavailable on the ESQ, additional data sources were sought, such as reports of outbreak investigations and details of epidemiological studies conducted, in order to deduce the reported date of exposure.

The incubation period was calculated as the difference in days between the reported date of symptom onset and the date of exposure. Where either or both dates were unknown, the case was excluded from the study.

6.2.5 Descriptive analysis and heterogeneity testing

Frequencies and percentages were calculated to describe the cases and outbreaks. Cases were described according to the reported characteristics including age, gender, ethnicity, geographical region of residence, infecting PT and reported symptoms. Outbreaks were described according to mode of transmission, PT and geographical region of outbreak.

In order to study the variability of incubation periods, I calculated the value of I^2 and performed pairwise Kolmogorov-Smirnov tests (KS tests) to compare the cumulative distributions of incubation periods between outbreaks as previously described in chapter 3.

Applying a bootstrap version of the function, the sampling was repeated 10,000 times to obtain a p-value for each pairwise comparison. Outbreaks containing one or two cases were excluded from the KS test in order for the comparison to be assessed using a significance level of 5%, since for these, it would be impossible for a p-value to fall below the 0.05 threshold used. I calculated the proportion of p-values below 0.05 and the probability of obtaining the observed proportion. A proportion sufficiently higher than the 5% expected by chance alone was considered to be evidence of heterogeneity. I used $p < 0.01$ in this test to indicate whether there was evidence that the proportion exceeded 5%.

6.2.6 Pattern of heterogeneity using hierarchical cluster analysis

Using the resulting p-values from the KS test, subgroups of outbreaks with limited evidence of heterogeneity, where the distribution of incubation period may be similar, were identified using hierarchical clustering analysis (101,102). Due to the small number of cases available per outbreak, I conducted a form of sensitivity analysis to examine the effect of smaller outbreaks on the subgrouping. The clustering analyses were conducted in stages starting with the large outbreaks with over 10 cases, and including smaller outbreaks in subsequent analyses.

6.2.7 Regression analyses

Using multi-level regression (MLR) modelling, I examined the relationship between the patient and outbreak characteristics and incubation period. The regression model included the patients' characteristics, the mode of transmission and region of outbreak as explanatory variables.

I first examined the association between incubation period and each variable in a univariate analysis. Where a significant association was observed, these variables were included in a multivariable model. A final model was developed using a backward step-wise process to exclude non-significant variables from the multivariable model. A likelihood ratio test was

used to compare each model to the null model which excluded the variable of interest and the significance level was set at a p-value of less than or equal to 0.1.

The univariate and multivariable models were fitted to log-transformed data to account for skewness. In order to account for the reporting of zero-day incubation period, which cannot be transformed to a log scale, a constant, c , was added to the value of the reported incubation period before transforming; where c is equal to 1.

In order to check for model sensitivity, I investigated alternative transforms on the data. The additional models evaluated were a linear MLR model using untransformed data and a generalised MLR model using the gamma family.

All analyses were conducted using the statistical software R version 3.2.3 (2015-12-10).

6.3 RESULTS

A total of 106 STEC outbreaks were reported in England and Wales during the study period of 2009 to 2016 affecting 1459 cases. The source of infection was unknown for 42 of the 106 outbreaks and these were excluded.

In two foodborne outbreaks, details of the investigation which included patient information and dates of exposure could not be retrieved, hence these outbreaks were excluded.

Reviewing the outbreak information available on the remaining 62 outbreaks and searching through the data reported by the affected 1013 cases, dates of exposure were identified for 205 cases in 41 outbreaks (Figure 18).

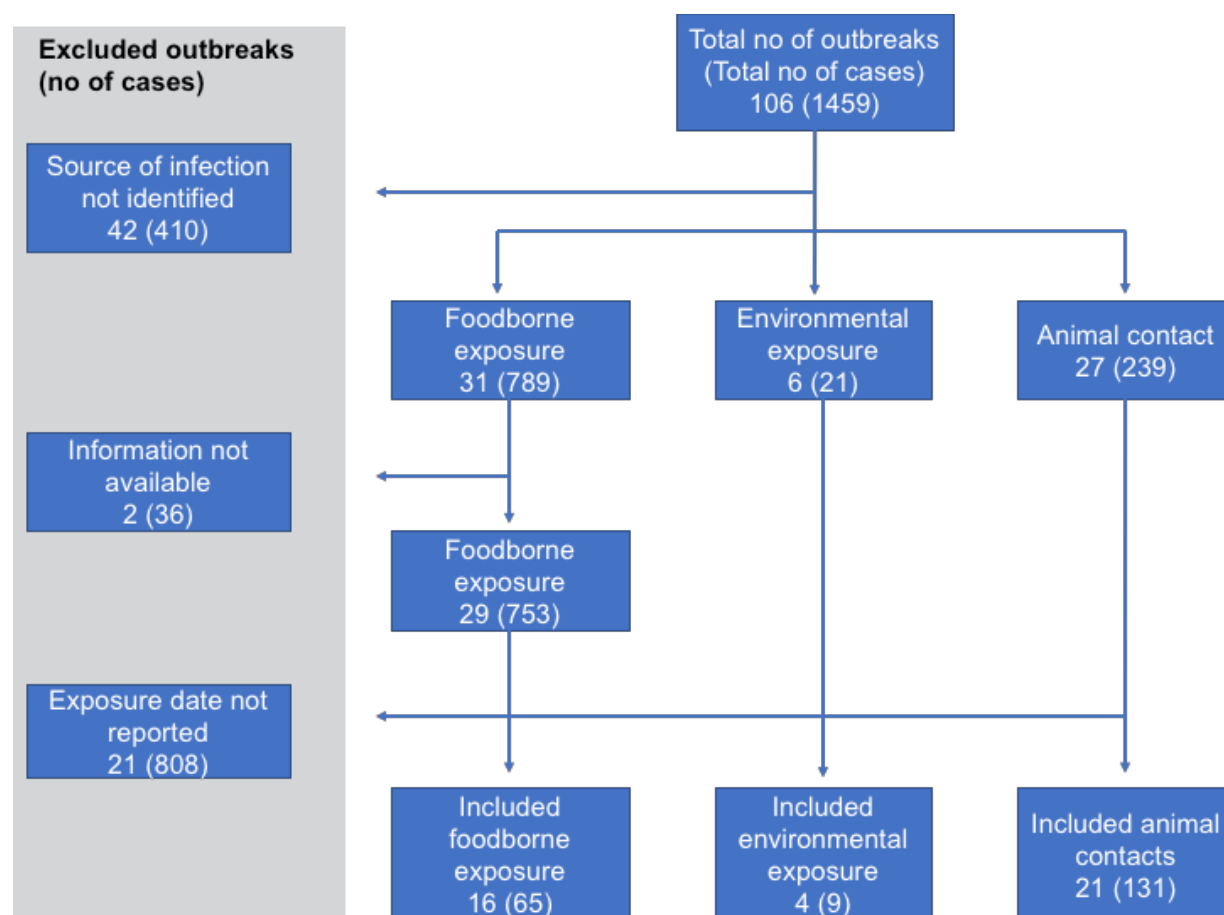


Figure 18 Flowchart showing selection of outbreaks and cases with available individual data in the individual patient data analysis of STEC cases

6.3.1 Comparing characteristics of excluded and included cases

The characteristics of cases excluded from the study were significantly different from those included in the study (Table 15). Distribution of cases according to age group (p-value <0.0001), ethnicity (p-value <0.0001), reported symptoms (p-value <0.05) and mode of transmission (p-value <0.0001) were significantly different in both groups of cases. All included cases were of serotype O157.

Table 15 Comparison of characteristics between excluded and included cases; and mean and median incubation period of cases included in the individual patient data analysis of STEC cases

Variable (Chi square for equality in proportion between excluded and included cases)	Number of excluded cases (%) (N = 1254)	Number of included cases (%) (N = 205)	Mean IP* for included cases	Median IP* for included cases
Age group (<0.0001)				
0 – 4 years	262 (20.9)	69 (33.7)	6.0	6.0
5 – 9 years	132 (10.5)	48 (23.4)	6.9	5.0
10 -19 years	140 (11.2)	23 (11.2)	5.1	4.0
20 - 29 years	120 (9.6)	26 (12.7)	5.2	3.5
30 - 39 years	219 (17.5)	16 (7.8)	5.9	4.0
40 - 59 years	169 (13.5)	10 (4.9)	3.2	2.0
60 - 79 years	161 (12.8)	10 (4.9)	4.0	3.0
80 and above	51 (4.1)	3 (1.4)	5.3	5.0
Gender (>0.5)				
Females	758 (60.4)	121 (59.0)	6.2	4.0
Males	496 (39.6)	84 (41.0)	5.5	4.0
Ethnicity (<0.0001)				
White	1147 (91.5)	166 (81.0)	5.9	4.0

Mixed ethnicity	9 (0.7)	3 (1.4)	6.0	5.0
Asian/Asian British	14 (1.1)	0		
Black/Black British	13 (1.0)	1 (0.5)	2.0	2.0
Chinese	2 (0.2)	1 (0.5)	7.0	7.0
Other	6 (0.5)	0		
Unknown	63 (5.0)	34 (16.6)		
Relevant occupation (>0.5)				
Food handler	60 (4.8)	8 (3.9)	5.2	4.5
Healthcare worker	75 (6.0)	4 (2.0)	5.5	4.5
Childcare workers	220 (17.5)	40 (19.5)	7.3	6.0
Geographical region of patient residence (<0.0001)				
East Midlands	78 (6.2)	4 (1.9)	3.7	3.0
East of England	121 (9.6)	17 (8.3)	8.7	5.0
London	96 (7.7)	32 (15.6)	7.5	6.5
North East	78 (6.2)	27 (13.2)	3.7	3.0
North West	217 (17.3)	27 (13.2)	4.6	4.0
South East	220 (17.5)	45 (21.9)	5.0	4.0
South West	214 (17.1)	16 (7.8)	5.1	4.0
Wales	25 (2.0)	1 (0.5)	6.0	6.0
West Midlands	95 (7.6)	9 (4.4)	9.3	8.0
Yorkshire and Humber	109 (8.7)	27 (13.2)	6.7	4.0
Unknown	17 (1.4)	0		
Phage Type				
PT 21/28	256 (20.4)	134 (65.4)	6.2	5.0
PT 8	543 (43.3)	18 (8.8)	4.4	3.0
PT 1	8 (0.6)	17 (8.3)	5.2	4.0
PT 2	87 (6.9)	14 (6.8)	3.7	3.0
PT 54	12 (1.0)	6 (2.9)	9.5	11.0

PT 32	72 (5.7)	5 (2.4)	4.6	4.0
PT 34	175 (14.0)	3 (1.5)	6.0	6.0
PT 4	14 (1.1)	3 (1.5)	3.7	5.0
PT 31	5 (0.4)	2 (1.0)	13.5	13.5
RDNC	11 (0.9)	2 (1.0)	5.5	5.5
PT 14	5 (0.4)	0		
PT 24	5 (0.4)	0		
PT 51	4 (0.3)	0		
PT 43	2 (0.2)	0		
PT 21	1 (0.1)	0		
PT 47	1 (0.1)	0		
Unknown	1 (0.1)	1 (0.5)		
Non-O157	52 (4.1)	0		
Reported symptoms (<0.05)				
HUS^	61 (4.9)	17 (8.3)	5.2	5.0
Diarrhoea	1012 (80.7)	191 (93.2)	5.8	4.0
Bloody diarrhoea	968 (77.2)	135 (65.9)	4.8	4.0
Nausea	900 (71.8)	102 (49.8)	5.6	4.0
Vomiting	904 (72.1)	87 (42.4)	5.2	4.0
Abdominal pain	993 (79.2)	164 (80.0)	5.6	4.0
Fever	874 (69.7)	57 (27.8)	5.3	4.0
Mode of transmission (<0.0001)				
Direct animal contact	108 (12.8)	131 (63.9)	6.9	5.0
Environmental exposure	12 (1.4)	9 (4.4)	3.6	3.0
Foodborne	724 (85.8)	65 (31.7)	4.1	3.0
Geographical region of outbreak (0.5)				
	Number of outbreaks (N = 95)	Number of outbreaks (N = 41)		
East Midlands	6 (6.3)	3 (7.3)		

East of England	3 (3.2)	3 (7.3)		
London	2 (2.1)	0		
National	18 (18.9)	4 (9.8)		
North East	7 (7.4)	2 (4.9)		
North West	15 (15.8)	7 (17.1)		
South East	13 (13.7)	5 (12.2)		
South West	13 (13.7)	7 (17.1)		
Wales	1 (1.1)	1 (2.4)		
West Midlands	3 (3.2)	1 (2.4)		
Yorkshire and Humber	11 (11.6)	8 (19.5)		
International	3 (3.2)	0		

* Incubation period; ^Haemolytic uraemic syndrome

6.3.2 Description of outbreaks

Of the 41 outbreaks included in the study, 51.2% (21/41) were associated with animal contact through contact with livestock during farm visits, 39% (16/41) were foodborne and 9.8% (4/41) were associated with environmental exposures, some of which included an outdoor sporting activity and animal fair. Of the 16 foodborne outbreaks, red meat or salad were the implicated food vehicles in 50% and 18.8% of outbreaks respectively (Table 16). Most of the outbreaks were localised, occurring in the Yorkshire and Humber (19.5%; 8/41), South West (17.1%; 7/41) and North West regions (17.1%; 7/41) of England. No outbreaks occurred in London only, however 84.3% (27/32) of the outbreak cases reported in London were associated with a particular outbreak in the South East. Four national outbreaks occurred, associated with nationally distributed products.

The most common PT causing outbreak was PT 21/28 accounting for 36.6% (15/41) of the outbreaks. PT 8 accounted for 17.1% (7/41) and there were four outbreaks with mixed PT's. The median incubation period of outbreaks ranged from two days, in an outbreak associated with environmental exposure, to 13.5 days in a foodborne outbreak (Figure 19).

Table 16 Characteristics of outbreaks included in the analysis in the individual patient data analysis of STEC cases

Variable	Number of outbreaks (N = 41)	Proportion
Mode of transmission		
Animal contact	21	51.2
Environmental exposure	4	9.8
Foodborne exposure	16	39.0
Geographical region of outbreak		
East Midlands	3	7.3
East of England	3	7.3
National	4	9.8
North East	2	4.9
North West	7	17.1
South East	5	12.2
South West	7	17.1
Wales	1	2.4
West Midlands	1	2.4
Yorkshire and Humber	8	19.5
Phage Type		
PT 21/28	15	36.6
PT 8	7	17.1
PT 2	5	12.2
PT 1	2	4.9
PT 32	2	4.9

PT 4	2	4.9
PT 54	1	2.4
PT 34	1	2.4
RDNC	1	2.4
Mixed PT outbreaks	4	9.8
Unknown	1	2.4

6.3.3 Heterogeneity between outbreaks

The calculated value of I^2 was 80.4% indicating high heterogeneity across the outbreaks. Twenty-four outbreaks were included in the KS test. Comparing the incubation period distribution between outbreaks showed evidence of heterogeneity. The proportion of p-values less than 0.05 was 15% and the probability of obtaining the observed proportion was <0.0001, confirming that the observed variability across outbreaks was unlikely to be solely due to chance.

6.3.4 Pattern of heterogeneity

The outputs of the hierarchical cluster analyses, showed that the outbreaks were split into two subgroups without evidence for heterogeneity. The large outbreaks with over 10 cases and those with over 5 cases formed the core subgroups mainly split into outbreaks with animal contact and foodborne outbreaks. As the smaller outbreaks were included, this pattern of subgrouping was still visible, however, with more variation as the small outbreaks gravitated around the larger outbreaks. This data-based subgrouping appears to be partly driven by the mode of transmission (Figure 20).

6.3.5 Description of cases

A total of 205 cases were included in the study, of which 57.1% (117/205) were children under ten years of age. Females accounted for 59% (121/205) of cases (Table 15).

Serotyping results were available for all cases, all of which were serotype O157. Phage typing results were available for 204 cases and PT 21/28 was most commonly reported (65.4%; 134/205) (Table 15).

Foodborne transmission accounted for 31.7% (65/205) of cases while 63.9% (131/205) of cases were attributed to direct animal contact. A large proportion of cases attributed to direct animal contact (49%; 64/131) were associated with one large outbreak (149).

Twenty one percent (45/205) of cases lived in the South East region of England, and of these, 80% (36/45) were cases associated with animal contact exposures. All but one of the cases living in London were associated with animal contact (96.9%; 31/32).

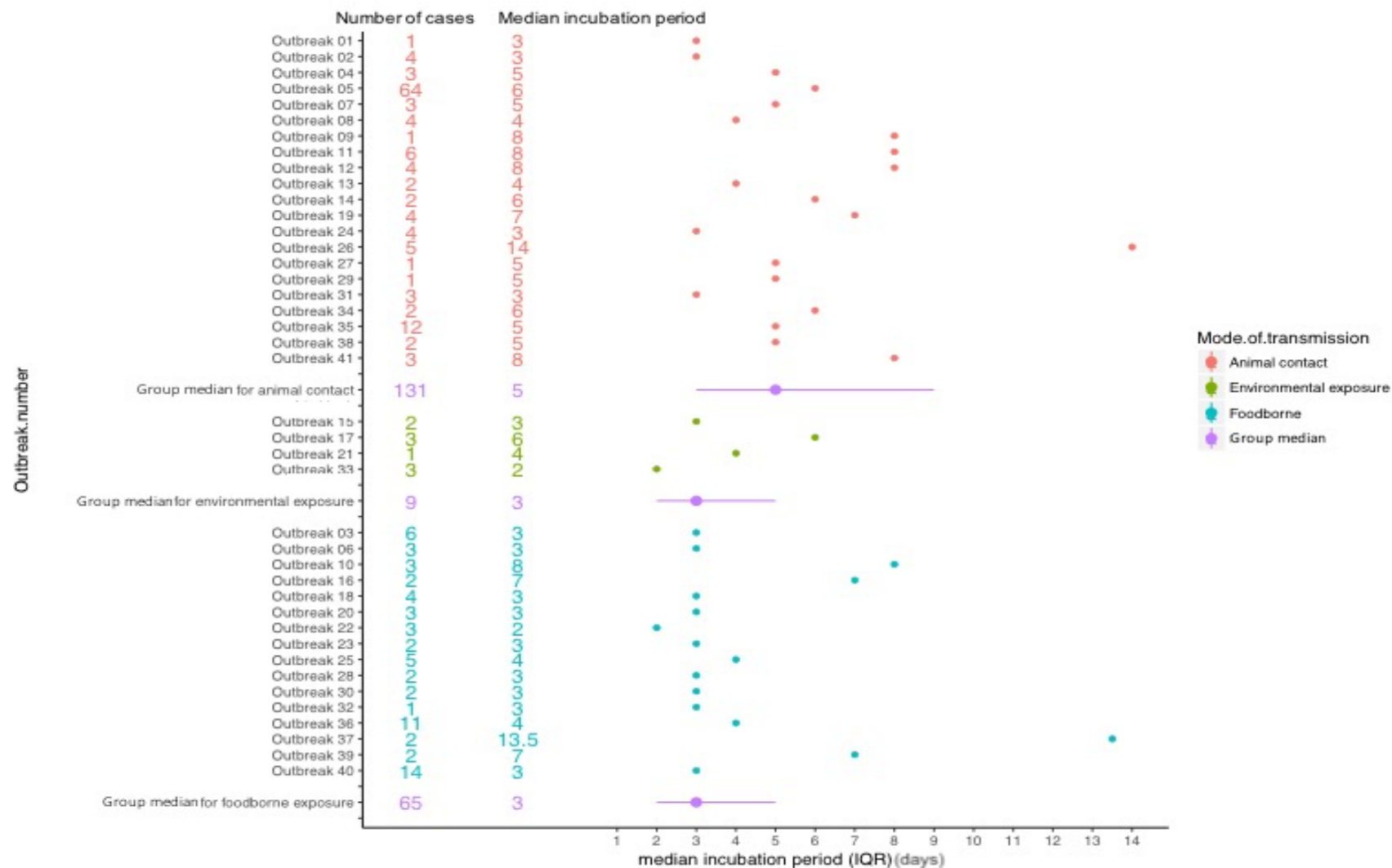


Figure 19 Forest plot showing median incubation period and IQR of outbreaks included in the individual patient data analysis of STEC case

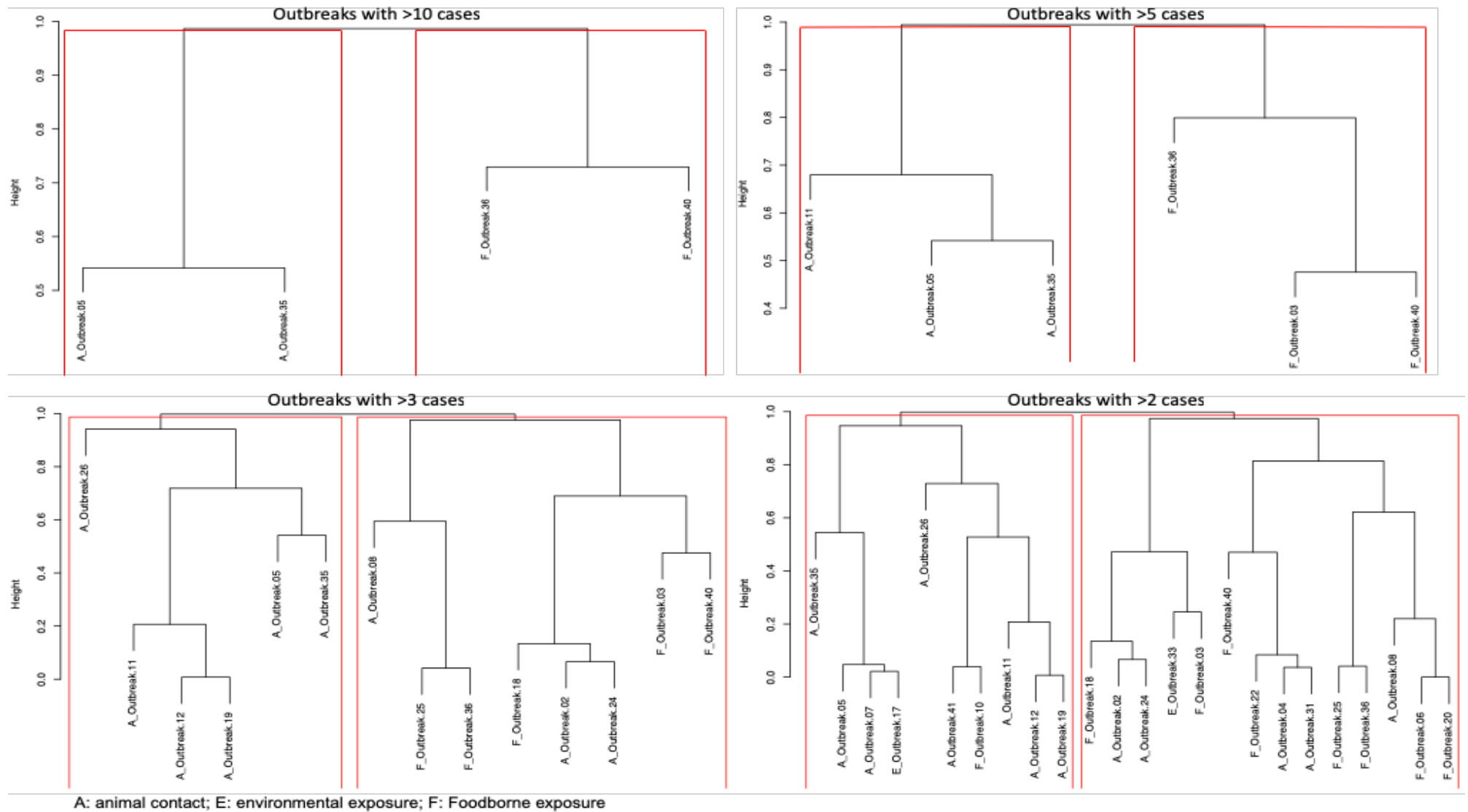


Figure 20 Dendrograms showing the clustering pattern driven by the mode of transmission in the individual patient data analysis of STEC cases

6.3.6 Reported symptoms

Cases reported a combination of symptoms and the most frequently reported symptoms were diarrhoea (93.2%), abdominal pain (80%), bloody diarrhoea (65.9%) and nausea (49.8%). The median time interval between the onset of illness and specific symptoms was zero days for diarrhoea, nausea, fever and abdominal pain and one day each for bloody diarrhoea and vomiting (Figure 21).

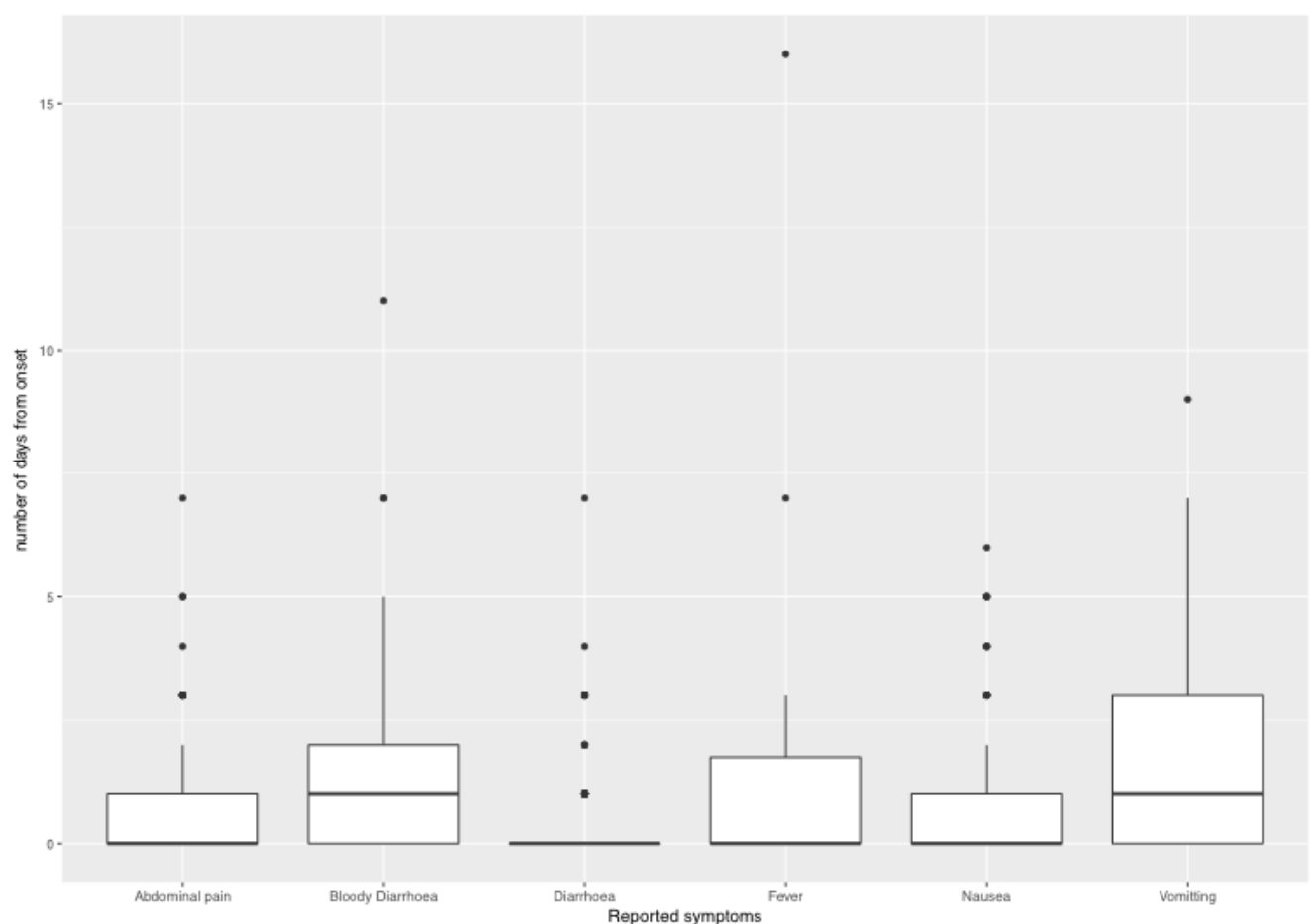


Figure 21 Boxplot showing median days and inter-quartile range between onset of first symptom and reported symptoms in the individual patient data analysis of STEC cases

6.3.7 Incubation period and associated factors

The median incubation period was four days with an interquartile range of three to seven days. Just over 10% (22/205) of cases reported an incubation period longer than 10 days (Figure 22). With the exception of one case of foodborne infection with an incubation period of 22 days, long incubation periods of eleven days or more were observed in 21 cases acquiring their infection from animal contact.

Using the log transformed data, the univariate analysis identified bloody diarrhoea and mode of transmission as variables that were significantly associated with the duration of incubation period (p-values of 0.001 and 0.01 respectively), while age group and vomiting showed a weak association (p-value of 0.1) but met the inclusion criteria to be included in the multivariable model (Table 17). In the final multivariate analysis, bloody diarrhoea and mode of transmission remained significantly associated with incubation period (p-values of 0.001 and 0.01 respectively). Cases reporting bloody diarrhoea had an incubation period of 1.19 days shorter (95% CI: -1.21 to -0.82) than cases without bloody diarrhoea. Compared to cases infected through environmental exposure, cases with illness due to direct animal contact had an incubation period of 2.7 days longer (95% CI: 0.20 to 9.79) (Table 17). From the sensitivity analysis, testing alternative link functions and residual distribution models had no effect on the results.

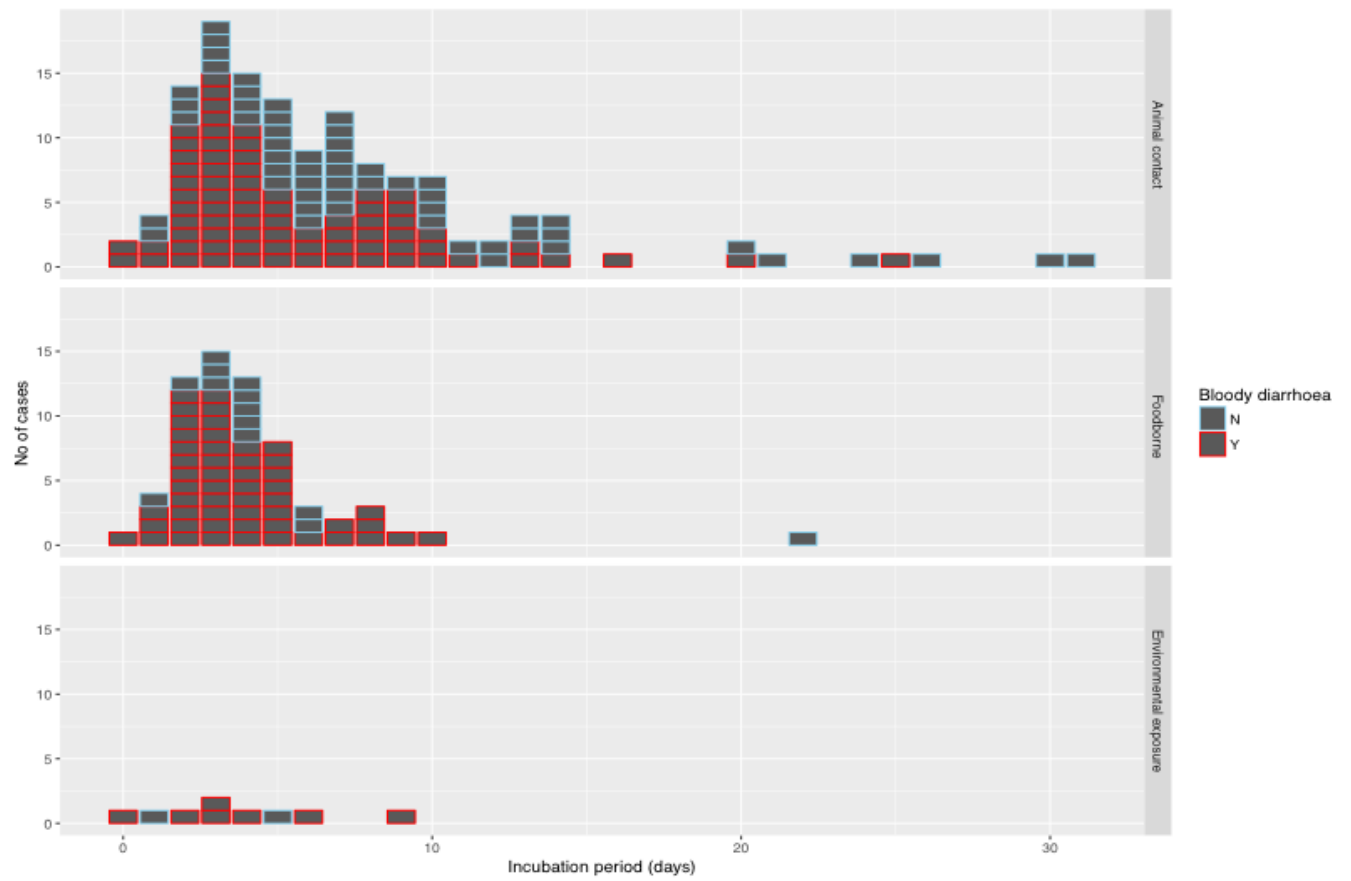


Figure 22 Histogram of reported incubation period by mode of transmission and indicating cases with bloody diarrhoea in the individual patient data analysis of STEC cases

Table 17 Multi-level regression model showing factors associated with incubation period in the individual patient data analysis of STEC cases

	Model one (Univariate model)		Model two (Full multivariate model)		Model three (Multivariate model excluding Age group)		Model four (Multivariate model excluding Vomiting)		Final model	
Variable	Difference in IP* (95% CI^)	p-value of LRT†	Difference in IP (95% CI)	p-value of LRT	Difference in IP (95% CI)	p-value of LRT	Difference in IP (95% CI)	p-value of LRT	Difference in IP (95% CI)	p-value of LRT
Age group										
0 – 4 years	Reference	0.1	Reference	0.6			Reference	0.6		
5 – 9 years	0.3 (-0.9, 1.8)		0.5 (-0.3, 2.9)				0.5 (-0.3, 2.8)			
10 -19 years	-0.8 (-2.1, 0.7)		-0.3 (-0.9, 1.9)				-0.2 (-0.9, 1.9)			
20 - 29 years	-0.6 (-1.9, 1.0)		0.07 (-0.8, 2.7)				0.1 (-0.7, 2.8)			
30 - 39 years	-0.6 (-2.1, 1.3)		-0.1 (-0.9, 2.9)				-0.1 (-0.9, 2.9)			
40 - 59 years	-2.5 (-3.6, -0.8)		-1.5 (-1.7, 0.4)				-1.4 (-1.7, 0.5)			
60 - 79 years	-1.4 (-2.9, 0.7)		-0.5 (-1.3, 2.8)				-0.4 (-1.2, 2.9)			
80 and above	0.8 (-2.4, 7.3)		0.4 (-1.5, 8.4)				0.5 (-1.4, 8.6)			
Bloody diarrhoea										
No	Reference	0.001	Reference	0.009	Reference	0.003	Reference	0.002	Reference	0.001
Yes	-1.6 (-2.1, -0.8)		-1.05 (-1.1, -0.4)		-1.1 (-1.2, -0.6)		-1.1 (-1.1, -0.7)		-1.1 (-1.2, -0.8)	
Vomiting										
No	Reference	0.1	Reference	0.4	Reference	0.5				

Yes	-0.6 (-1.1, 0.3)		-0.3 (-0.6, 0.8)		-0.2 (-0.6, 0.9)					
Mode of transmission										
Environmental exposure	Reference	0.01	Reference	0.08	Reference	0.01	Reference	0.09	Reference	0.01
Direct animal contact	2.2 (0.1, 8.4)		2.3 (0.1, 9.6)		2.7 (0.2, 9.9)		2.3 (0.05, 9.5)		2.6 (0.2, 9.7)	
Foodborne	0.6 (-0.6, 4.6)		1.1 (-0.6, 6.2)		0.8 (-0.7, 5.4)		1.1 (-0.6, 6.1)		0.8 (-0.7, 5.4)	
Gender										
Females	Reference	0.6								
Males	-0.2 (-0.3, -0.0)									
Geographical region of outbreak										
East Midlands	Reference	0.8								
East of England	-2.3 (-2.9, 3.1)									
National	-1.1 (-2.5, 6.1)									
North East	-2.2 (-2.9, 3.1)									
North West	-2.7 (-2.9, 0.7)									
South East	-2.1 (-2.6, 2.6)									
South West	-1.6 (-2.6, 4.1)									
Wales	-1.1 (-3.4, 23.3)									
West Midlands	-3.1 (-3.5, 6.0)									
Yorkshire and Humber	-1.2 (-2.3, 4.5)									

Phage Type										
PT 1	Reference	0.3								
PT 2	-1.1 (-1.9, 2.2)									
PT 21/28	-0.2 (-1.3, 3.3)									
PT 31	7.1 (-0.2, 38.7)									
PT 32	0.01 (-1.8, 7.4)									
PT 34	1.1 (-1.7, 13.7)									
PT 4	-1.1 (-2.3, 5.6)									
PT 54	3.2 (-0.5, 16.4)									
PT 8	-0.3 (-1.5, 3.8)									
RDNC	1.3 (-1.8, 16.1)									
HUS										
No	Reference	0.3								
Yes	-0.7 (-1.6, 1.0)									
Diarrhoea										
No	Reference	0.8								
Yes	-0.1 (-1.1, 2.6)									
Nausea										
No	Reference	0.2								
Yes	-0.4 (-1.1, 0.5)									
Abdominal pain										

No	Reference	0.3								
Yes	-0.5 (-1.2, 0.8)									
Fever										
No	Reference	0.6								
Yes	0.2 (-0.9, 0.9)									

*Incubation period; †p-value of likelihood ratio test; ^95% confidence interval

6.4 DISCUSSION

The mean and median incubation period observed in this study were six days and four days respectively with an interquartile range of 3 – 7 days, within the range reported in the systematic review in chapter 5 where the means of subgroups ranged between 3.5 and 8.1 days. From the I^2 and KS tests, there was strong evidence for heterogeneity amongst outbreaks included in the study. Two clusters were identified based on similar incubation period distributions without remaining evidence for heterogeneity within clusters. One was dominated by foodborne outbreaks and one by those arising from direct animal contact, although the separation was not complete (Figure 20). Consistent with this, the forest plot (Figure 19) shows foodborne outbreaks having shorter incubation periods (median 3 days) and outbreaks attributed to animal contact longer (median 5 days), although again with exceptions. One large outbreak with incubation time data on 64 cases (149) contributed substantially to the animal contact data, and with median of 6 days, it was longer than the overall median of 5 days across all 131 animal contact associated cases.

The regression analyses showed that cases reporting bloody diarrhoea had shorter incubation times than cases not reporting this symptom. This finding was dependent on inclusion of the data from the 64 people in the largest animal contact outbreak (Outbreak 05 in Figure 19), in which 46% of cases reported bloody diarrhoea compared to 74% from the other outbreaks in the study and the significant association was lost in a sensitivity analysis excluding that outbreak.

Haemorrhagic colitis (HC) and the associated symptoms of bloody diarrhoea and abdominal pain is an afebrile syndrome of acute STEC infection (171). The ability of the bacteria to adhere to the intestinal mucosa and produce cytotoxins contributes to its virulence factors (172). Although bloody diarrhoea is itself among the later symptoms reported by the cases, it can be suggested that the virulence factors of the bacteria may be associated with both

bloody diarrhoea and the early pathogenesis of the disease leading to earlier onset of other symptoms.

There are several possible explanations for the shorter median incubation period identified among cases resulting from foodborne transmission compared to direct animal contact. The quantity of pathogen ingested (infecting dose), an important determinant of incubation period (65,136,173), could vary across transmission route, with for example, higher ingested doses from contaminated food than direct animal contact. The food matrix may also be important. Fatty foods have been known to act as a buffer for some gastrointestinal bacteria (129), protecting it from gastric acid, therefore increasing the likelihood of infection occurring. Although STEC can tolerate gastric acid (174), ingesting it with fatty foods probably increases its survival chances and may speed up the infection process. Ingesting bacteria with no accompanying food may decrease chances of survival through the stomach. Other food associated effects are also possible. In a mouse experiment, high fibre diets were reported to enhance the presence of Shiga-toxin receptors leading to higher colonization of STEC (175).

There was no association of incubation time with pathogen characteristics (PT), age or development of HUS. In England and Wales, serotype O157 is most commonly associated with human disease (141), and also commonly implicated in outbreaks (147). However, there is a diagnostic bias for detection of O157 making it more likely to be investigated than other STEC serotypes. This may account for the difference between excluded and included cases where all non-O157 cases were excluded from the study as information on time of exposure was unavailable. I cannot therefore comment on differing incubation periods across serotypes, with some published evidence that e.g. serotype O104 infections may have a particularly long incubation period (57). According to Byrne et al (141), the three most commonly reported STEC O157 phage types in England as at 2012 were PT21/28, PT8, and PT32. Similarly, in the study population, PT21/28 and PT8 were commonly reported, in addition to PT1 and PT2. No association was observed between phage type and incubation

period. With the use of genomics in the characterisation of STEC (176), additional microbiological data are available for future analysis and it may be possible to identify pathogen factors affecting incubation period.

Age was not associated with altered incubation period in this data. This contrasts with other reports where children have reported shorter incubation periods (177–179). Because animal contact was a more common source among children, I considered that there might be confounding factors hiding the age effect. However, the results were consistent even after adjustment for source of infection.

Haemolytic uraemic syndrome is a known complication of STEC O157 infections. This study, with 17 of 205 cases reporting HUS having a mean incubation period of 5.1 days, found no significant association with incubation period. A consistent result was reported in the systematic review in chapter 5, where no significant difference was observed in the incubation period of outbreaks involving HUS cases and those that did not.

This study has several limitations. The cases analysed were a small total of all of the cases in the included outbreaks due to missing information. In contrast with cases associated with animal contact, the incubation period could not be calculated for 86% of foodborne associated outbreaks which had to be excluded. The impact on internal validity is likely to be limited unless those excluded had substantially different incubation periods or different associations between risk factors and incubation period than those included. There were a few cases associated with environmental exposure so that I was not able to offer any robust inference on the distribution of incubation period associated with this transmission route. Additionally, the proportion of symptoms reported by cases was significantly different between included and excluded cases with a higher proportion of bloody diarrhoea being reported amongst the excluded cases. If these excluded cases had systematically different incubation periods, this could result in bias in the estimates of the distribution of incubation periods.

6.5 CONCLUSIONS

In conclusion, I used individual patient data, collected across outbreaks, to estimate the incubation period distribution of STEC. I analysed patient characteristics like age, gender and reported symptoms as well as outbreak characteristics such as mode of transmission and geographical location in order to identify factors affecting the distribution of incubation period. The median incubation period was four days and identified associated factors were mode of transmission and developing bloody diarrhoea. Public health professionals should therefore consider the effect of patient and outbreak characteristics on the incubation period, and outbreak investigations should be tailored accordingly to either extend or restrict the investigative period when collecting exposure details. Further research may be required to determine the effect of characteristics that were not examined in this study such as STEC serogroup and genomic data, underlying medical conditions and ongoing medications.

A major benefit to analysing individual patient data from outbreaks was having patient level information that might influence the distribution of incubation period. However, there were several limitations to the quality of recorded data to support the purposes as might be expected from systems not established for this particular purpose. The detail of information available for analysis depended on how comprehensively the cases answered the questions on the ESQ or how efficiently they were asked by the interviewer. I observed that the question regarding the date of exposure was mostly unanswered, as well as not being explicitly included for every exposure. It was often easier to obtain dates of exposure from cases in outbreaks associated with a location such as farm visits, than it was to obtain dates of exposure from outbreaks associated with purchased food items. For purchased food items, the ESQ records the date the item was purchased and not the date of consumption. Similarly, other patient information which may be useful for this study was not collected at all, such as underlying medical conditions and ongoing medications. Studies such as this one that show the research potential of routine surveillance data may guide future

surveillance to ensure that such applied research outcomes prospectively in planning surveillance is considered. Consideration of research applications is likely to inform content and form of data capture as well as consent processes where relevant.

CHAPTER 7: INCUBATION PERIOD OF TYPHOIDAL SALMONELLOSIS: A SYSTEMATIC REVIEW AND META-ANALYSIS

ABSTRACT

Salmonella Typhi is a human pathogen that causes typhoid fever. It is a major cause of morbidity and mortality in developing countries and is responsible for several outbreaks in developed countries. Studying certain parameters of the pathogen, such as the incubation period, provides a better understanding of its pathophysiology and its characteristics within a population. Outbreak investigations and human experimental studies provide an avenue to study these relevant parameters.

In this chapter, I conducted a systematic review of outbreak investigation reports and experimental studies, extracted reported data, tested for heterogeneity, identified subgroups of studies with limited evidence of heterogeneity between them and identified factors that may contribute to the distribution of incubation period.

Following identification of relevant studies, both raw and summary incubation data were extracted. The data was tested for heterogeneity by deriving the value of I^2 and conducting a KS-test to compare the distribution between studies. I performed a linear regression analysis to identify the factors associated with incubation period and using the resulting p-values from the KS-test, conducted a hierarchical cluster analysis to classify studies with limited evidence of heterogeneity into subgroups.

Thirteen studies were identified to be included in the review and raw incubation period data was extracted from eleven. The value of I^2 was 84% and the proportion of KS test p-values that were less than 0.05 was 63.6% indicating high heterogeneity not due to chance.

Regression analysis identified vaccine history and attack rates as factors that may be associated with incubation period, although these were not significant in the multivariable

analysis (p-value: 0.1). From the hierarchical clustering analysis, studies were classified into five subgroups. The mean incubation period of the subgroups ranged from 9.7 days to 21.2 days. Outbreaks reporting cases with previous vaccination history were clustered in a single subgroup and reported the longest incubation period.

7.1 INTRODUCTION

Typhoid fever is a systemic infection caused by the bacterium *Salmonella enterica* subsp. *enterica* serotype Typhi (S. Typhi). It has been a major human pathogen for thousands of years, and while incidence has greatly reduced in developed countries, it remains a major cause of morbidity and mortality in developing countries (180). Laboratory diagnosis of typhoid fever involves culture and identification through biochemical and serological means (181). The mode of transmission is faecal-oral via ingestion of contaminated food and water or contact with an infected carrier (63). Although it is a gastrointestinal infection, diarrhoea and vomiting are atypical symptoms of typhoid fever. People suffering from typhoid fever typically complain of systemic symptoms including gradual onset of fever, general malaise, headaches and abdominal pain (182). Prevention of typhoid fever, like most other gastrointestinal infections, includes good hand hygiene (183) and exclusion of infected food handlers. In addition, typhoid vaccine has been known to offer some protection (184).

Although typhoid fever in developed countries is mostly a sporadic disease associated with travel to endemic countries, outbreaks are still commonly reported (185,186). Factors such as improper cooling, inadequate heating of food, contact with contaminated raw products and infected food handlers have been implicated as contributory factors of outbreaks (41).

Investigation of outbreaks contributes to the control and reduction of the burden of disease by identifying and eliminating the source of infection. Information obtained from observational studies of outbreak investigations, particularly outbreaks with well-defined point of exposures, as well as experimental studies provide a means of studying and understanding the disease by studying certain parameters in real-life and under laboratory conditions.

Incubation period of typhoid fever, which is the time between exposure and onset of clinical symptoms, is one of the essential parameters to be studied. It is expected to be a

distribution rather than a single estimate as a result of individual variation based on factors such as dose response, food matrix and host characteristics. An accurate knowledge of the incubation period distribution of typhoid fever is relevant for surveillance and implementation of public health interventions. It also contributes towards effective outbreak investigation as incorrect estimations can be misleading if they were used to determine the time of exposure (187). Accurate knowledge of the incubation period helps in correctly classifying primary and secondary cases and exclusion of travel related cases. The incubation period also offers insights into the pathophysiology of typhoid fever and is important in conducting epidemiological and ecological studies (54).

Despite the importance of knowing the accurate incubation period distribution of typhoid fever, large public health organisations give a wide range with no clear indication of which durations are common and which are rare. According to the World Health Organisation (WHO), incubation periods ranging from three to sixty days have been reported (63). The Centres for Disease Control and Prevention (CDC) reported the incubation period as three to thirty days (188).

In this chapter, I have systematically reviewed the literature for outbreaks of well-defined point source exposures and human experimental studies of *S. Typhi* with the aim of estimating the distribution of the incubation period and identify factors that may explain any variation observed. I extracted summary estimates and individual incubation period data reported, tested for the presence of heterogeneity, identified factors that may contribute to heterogeneity and defined subgroups of studies that can be combined for analysis.

7.2 METHODS

7.2.1 Research questions and modified PICO elements

The questions I aimed to answer in this review were:

1. What is the distribution of incubation period of *Salmonella* Typhi in humans?
2. What factors influence the incubation period?

Population studies /Participants -	Laboratory confirmed cases of <i>Salmonella</i> Typhi in a point source exposure outbreak or continuous source outbreaks where dates of exposure and onset are known for each case or experimental study. Probable cases of typhoidal salmonellosis based on clinical presentation and case definitions in the context of a point source exposure outbreak.
Infectious agent -	<i>Salmonella enterica</i> subsp. enterica serotype Typhi
Comparator -	Host factors and any other factors such as vaccine history, ingested dose, food vehicle, case definitions
Outcome -	Time from exposure to onset of clinical illness as described or defined by the authors including fever, abdominal pain, vomiting etc.

7.2.2 Literature search

I carried out a systematic literature search of peer reviewed publications on PubMed to identify observational studies and experimental studies reporting incubation period.

Combining all the terms of interest, the compound search string was formed: *Salmonella* Typhi AND Humans AND (Outbreak* OR Experiment*). The reference list of review papers was also screened to identify other relevant studies that may have been missed in the original search. The search was carried out between 29 May to 24 June 2017.

7.2.3 Selection process and quality assessment

Each article went through a rigorous selection process, and relevant articles were assessed for the quality of reported data. The selection and assessment processes were done in the following phases:

- 1) Screening of titles and abstracts for articles reporting typhoidal salmonellosis
- 2) Screening of full texts for reporting of incubation period
- 3) Review of full texts to assess exposure times and quality of reported data according to defined quality assessment criteria.

The quality assessment criteria used in this review focused on assessing the quality of the incubation period data reported and not the quality of the overall study (Table 1). The process involved two reviewers independently assessing each study and comparing results. Where there was a difference in opinions, discussions were held until a consensus was reached.

7.2.4 Data extraction process

Relevant data was extracted using a pre-developed pro forma (Table 18). The type of data extracted from all studies included general information on the article, characteristics of the observational study or experimental study, details of the organism, attack rate, setting of exposure, details of case definition and summary measures of the incubation period such as mean, median, mode and range where available.

Some studies further reported individual incubation data either as an epidemic curve or descriptive table. Where an epidemic curve was provided, I used a free online data extraction tool called WebPlotDigitizer version 3.10 (96) to extract the individual incubation data. If data was provided as a descriptive summary table, individual incubation data was equally extracted.

The unit of measurement reported and extracted was in days.

Table 18 Details of data to be extracted in the review of *S. Typhi*

Section	Information to be collected
General information	- Year of publication - Title of article - Authors - Type of publication (journals, conference abstract, grey literature, etc.) - PubMed ID (where applicable)
Study characteristics	- Year of study - Study design (cohort, case-control, experimental, case series) - Country of study - Age distribution - Comments on method or quality of study
Pathogen characteristics	- Infectious agent - Species - Subtype
Outcome data/ results	- Case definition - Reported incubation period (individual data, mean, median mode and range) - Derived or calculated summary estimates incubation period (raw data extracted) - Source of calculated data (epidemic curve or author description) - Incubation period to particular symptoms
Factors that could affect incubation period	- No of exposed cases - No of people affected

	- Setting - Mode of transmission - Food vehicle (for foodborne infections only) - Patient characteristics (e.g. age, previous infection or treatment, underlying illness, vaccine history)
Any other relevant information	- Any other relevant information

7.2.5 Descriptive analyses

I calculated frequencies and percentages, summarising the characteristics of all studies.

Characteristics summarised included: study design (observational or experimental), study type for observational studies (case-control, cohort or descriptive), year of study, country of study, age distribution of cases, mode of transmission and setting of exposure.

Using the extracted individual incubation data, I re-created the epidemic curves of the studies plotting each graph on a standard x-axis indicating incubation period ranging from zero to forty-five days and individual y-axis indicating number of cases in the outbreak or experimental study.

I compared the mean incubation period between the observational studies and experimental studies using a t-test to confirm if there was any statistical difference.

7.2.6 Statistical analyses

Using the individual incubation data, I tested for the presence of heterogeneity and defined the pattern of heterogeneity. Using both the individual and summary incubation data, I identified factors that may explain heterogeneity. Analyses were carried out using the statistical software R version 3.2.3 (2015-12-10) – “Wooden Christmas Tree” (127).

7.2.6.1 Testing for heterogeneity

I tested for heterogeneity across the studies in two ways. First, I calculated the value of I^2 by deriving the Friedman's statistic and inputting it into the formula: $I^2 = 100\% \times (\frac{Q-df}{Q})$; where Q is the Cochran's heterogeneity statistic, but in this case, the Friedman's statistic, and df is the corresponding degree of freedom. A p-value of less than 0.05 from the Friedman's test provided evidence of heterogeneity between the studies. Furthermore, the value of I^2 was interpreted according to the Cochran suggested threshold (99) to determine the magnitude of heterogeneity.

The second way I tested for heterogeneity was by performing a two-sample Kolmogorov-Smirnov test (KS test) which compares the cumulative distributions between the studies. I applied a bootstrapped version of the function, repeating the sampling 10,000 times in order to derive p-values that will provide improved coverage due to potential ties in the data comparisons.

I further compared the p-values to confirm if any observed heterogeneity was due to chance. I calculated the proportion of p-values below 0.05 and the probability of obtaining the observed proportion. Statistical evidence of heterogeneity in the reported incubation period was available if the probability was less than 0.01.

7.2.6.2 Identifying factors that explain heterogeneity

I performed a linear regression analysis using the summary incubation data available from all studies. The data was fitted to a generalised linear model with gamma as the family function to account for skewness of the data and the link function used was 'identity'. The effect of the explanatory variables on the mean incubation period was examined using a univariate model, and where there was a significant association (p-value 0.05), the associated variables were included in a multivariable model to test for confounding.

7.2.6.3 Identifying subgroups of studies/Describing pattern of heterogeneity

Once heterogeneity was confirmed, the next step was to identify subgroups of studies with limited evidence of heterogeneity. Hierarchical clustering analysis was used to statistically identify subgroups of studies. Studies that reported only summary data (studies 7 and 11) or aggregated raw data (studies 1 and 3) were excluded from the cluster analysis.

The p-value of the KS test was converted to a dissimilarity matrix and used to create a hierarchical cluster to show a graphical representation of the dissimilarities between the studies. The cluster analysis algorithm used was the complete linkage method. The output was a dendrogram showing a compact visualisation of the dissimilarity matrix.

In order to identify clusters within the dendrogram, a cut-off point at which the studies were significantly allocated to subgroups had to be determined. Considering that the KS test was a pairwise test, thereby increasing the likelihood of a type 1 error (observing one significant result due to chance), a pragmatic adjustment to the significance level (0.05) was applied by dividing it by the number of studies included in the KS test. I then subtracted the corrected p-value from one to derive a cut-off point from which studies with limited evidence of heterogeneity can be defined within subgroups.

7.2.7 Subgroup analysis

Individual incubation period data were pooled to generate a dataset for each subgroup.

Summary statistics and outcome measures were derived including:

- Number of studies included in a subgroup
- Total number of cases (sum of cases in all studies included in a subgroup)
- Mean and median incubation period of cases within a subgroup
- Variance, skew and kurtosis of incubation period of cases within a subgroup
- Mean attack rate of studies within a subgroup
- Reported vaccination history of studies within a subgroup

A forest plot showing the distribution of the mean incubation period and the corresponding 95% confidence interval was created. The reported mean incubation period was used to allocate studies without individual patient data to appropriate subgroups for illustration.

7.2.8 Risk of Bias

The data was analysed for the presence of small-study effect using a funnel plot to visualise the relationship between sample size and incubation period.

7.2.9 Determining the distribution of incubation period data

The raw incubation period data extracted were aggregated, and a Cullen and Frey plot was created to select the distributions that best fit the data.

7.3 RESULTS

7.3.1 Literature search and selection process

A total of 510 articles were retrieved following the search in PubMed. Titles and abstracts were screened to identify relevant articles. Articles were excluded if they were non-human related studies, or non-typhi studies. Studies that were not reported in English were also excluded. This resulted in 180 articles available for full text screening of incubation period data after excluding 330 irrelevant articles. Searching through the reference list of other review papers, thirteen articles that were not included in the original search were identified, bringing the number of articles available for full text screening to 193. Excluding articles that did not report incubation period resulted in 86 articles that went through the quality assessment process. Seventy-three articles did not meet the quality assessment criteria (Appendix 10) and were further excluded (Figure 23). The resulting number of articles available for inclusion in the study was thirteen (Appendix 11), all of which reported individual incubation data with the exception of studies 7 and 11.

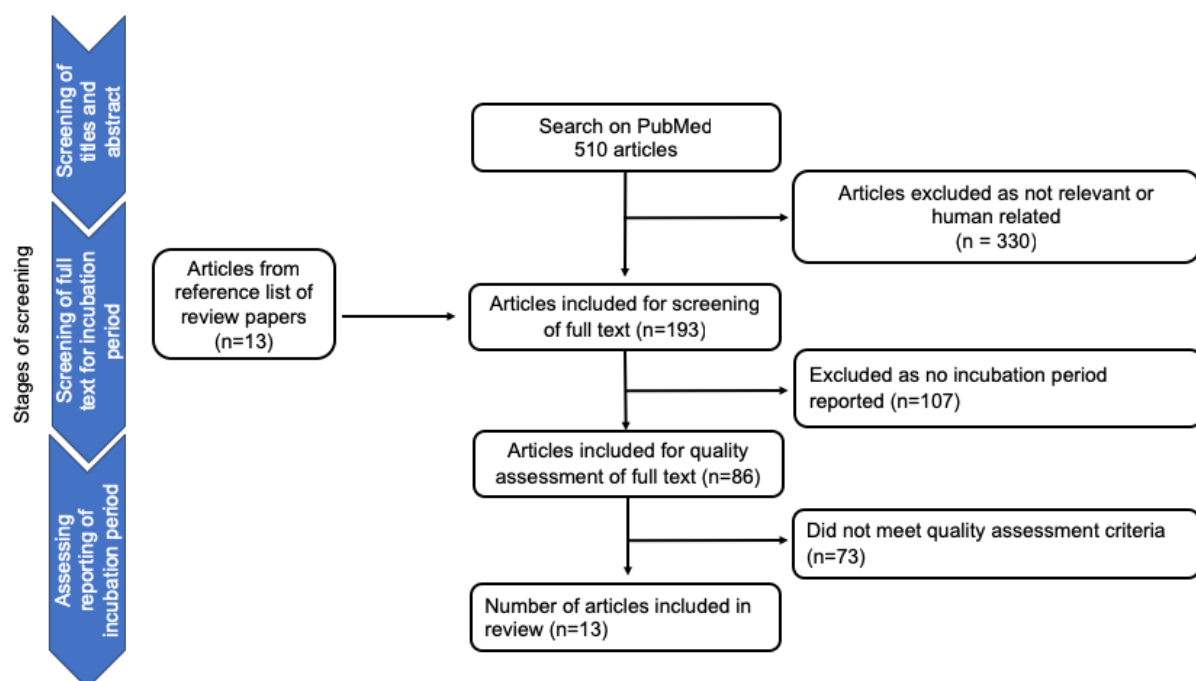


Figure 23 Flowchart of study selection process in the review of incubation period of *S. Typhi*

7.3.2 Descriptive analysis of studies

All studies included in this review reported incubation period of cases associated with *S. Typhi*, however, very few studies reported the microbiological characteristics (phage types) of the infecting organism.

Studies included in this review were published between 1914 and 2016. Sixty one percent (8/13) of the outbreaks or experiments took place before year 2000 with the majority (5/13; 38.5%) occurring in North America (Table 19). Two studies were experimental and the remainder were epidemiological studies including descriptive studies (4/13; 30.8%), case control studies (4/13; 30.8%) and retrospective cohort studies (3/13; 23.1%). The eleven outbreaks with individual patient data resulted in a total of 635 cases.

The settings of exposure varied and the most common reported setting was private events with catered meals (5/13; 38.5%). Outdoor activities and picnic events were reported in two outbreaks each (2/13; 15.4%). With the exception of one outbreak that was caused by exposure to surface water at a recreational park, all infections were foodborne and the most frequently reported food vehicles were salad and vegetables (3/13; 23.1%), red meat and dairy both accounting for two outbreaks each (2/13; 15.4%). The food vehicle for two outbreaks was unknown. (Table 19).

The funnel plot did not indicate any effect of study size on the reported incubation period as the data points were evenly distributed around the x-axis (Figure 24).

Reviewing the re-created epidemic curves from the eleven studies that provided individual incubation data, I observed a range of distributions in the plots (Figure 25). The mean incubation period ranged from seven days to 21.4 days (Table 20). The minimum incubation period reported was two days and the maximum was 41 days.

From the t-test conducted, the mean incubation period of the observational studies differed statically from the mean incubation period of the experimental studies (p-value 0.03).

Table 19 Characteristics of included studies in the review of incubation period of *S. Typhi*

	Number	Proportion
Total number of studies	13	
Year of study		
Before year 2000	8	61.5
2000 and later	4	30.8
unknown	1	7.7
Region of study		
Europe	3	23.1
North America	5	38.5
Africa	2	15.4
Asia	2	15.4
Unknown	1	7.7
Study design		
Descriptive	4	30.8
Case control	4	30.8
Retrospective cohort	3	23.1
Experimental study	2	15.4
Setting of exposure		
Catered meal	5	38.5
Outdoor activity	2	15.4
Picnic	2	15.4
Experimental study	2	15.4
Community	1	7.7
Restaurants	1	7.7
Food vehicle category		

Red meat	2	15.4
Dairy	2	15.4
Salad	3	23.1
Non-foodborne	1	7.7
Other	3	23.1
Unknown	2	15.4

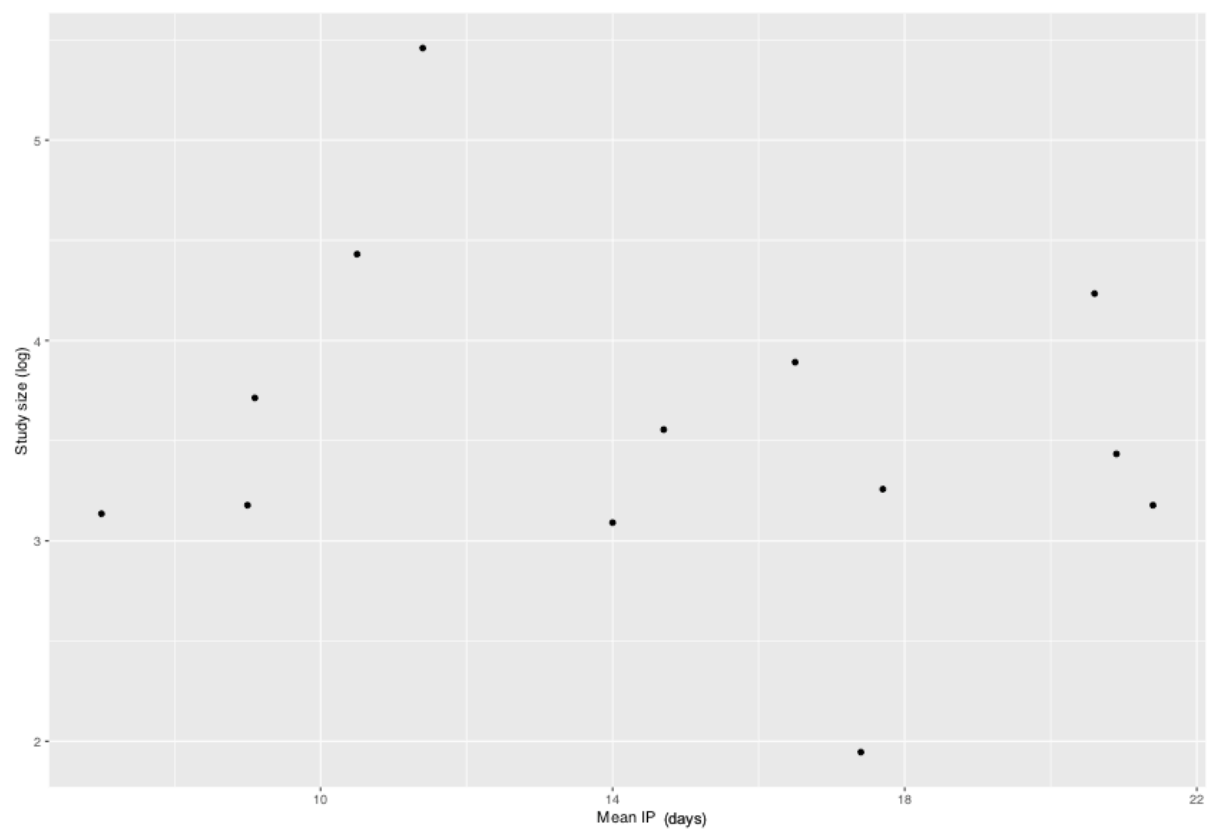


Figure 24 Funnel plot testing small study effect in the review of the incubation period of *S. Typhi*

Identification of cases was dependent on the case definition given by the authors of each study. In eight studies, microbiological confirmation was used to define cases, in two studies (studies 6 and 13), a very specific case definition including epidemiological link was used and in three studies, the case definition was less specific but also including epidemiological link (Table 20).

Of the twelve studies where the attack rates could be calculated, five had an attack rate higher than 50% and the highest attack rate was 79.31% in an outbreak involving ingestion of contaminated ice cream (189). This outbreak also recorded the second highest case fatality rate.

Five studies reported vaccination history of cases, and of these, two reported previous vaccination of some or all of their cases. These two studies reported the longest mean incubation period of 21.4 and 20.9 days. The attack rates in these populations were below 50%, however, these were not the lowest attack rates reported (Table 20).

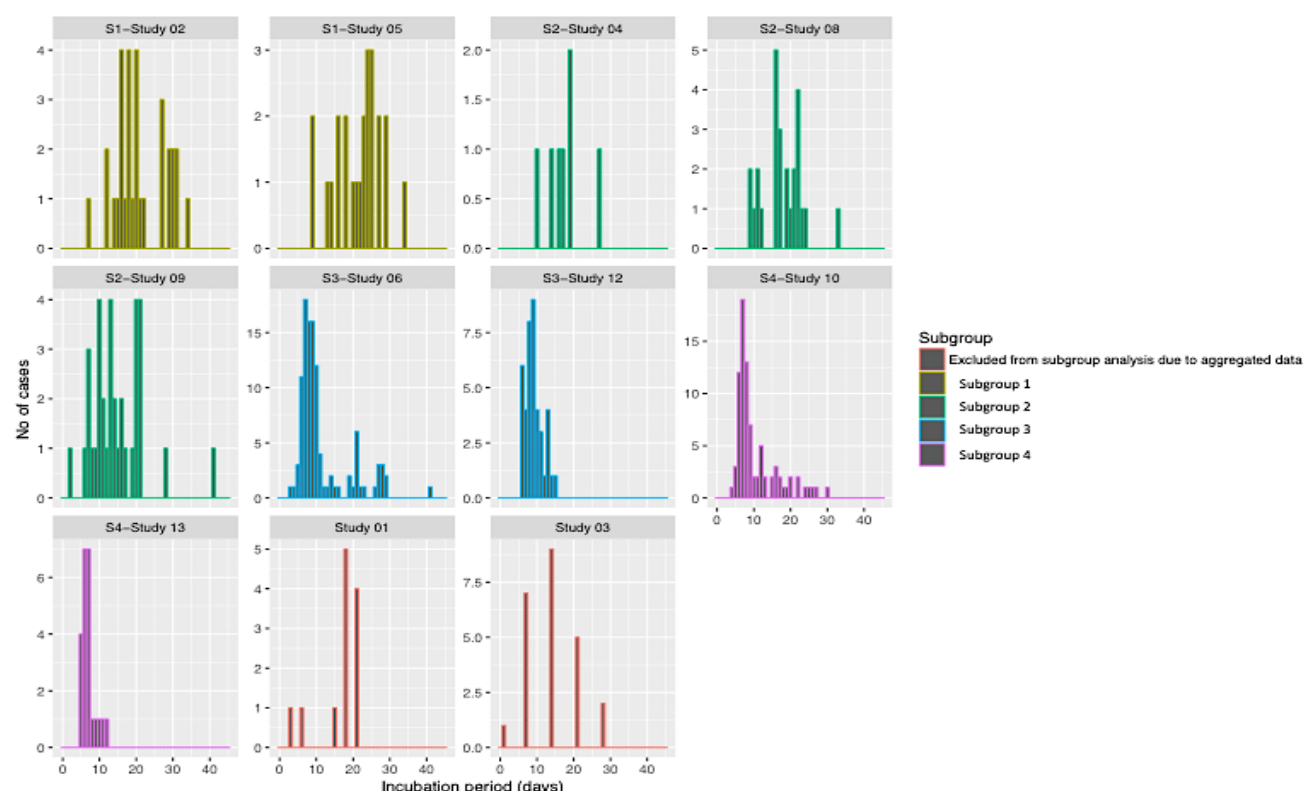


Figure 25 Re-created epidemic curves of eleven studies in the review of the incubation period of *S. Typhi*

Table 20 Additional characteristics of included studies in the review of incubation period of *S. Typhi*

Study number	Mean incubation period	No. affected (attack rate)	No. hospitalised (deaths)	Age Distribution	Source	Setting of exposure	Food vehicle	Ingested dose	Vaccine history	Case definitions as described by the studies
Study 1	16.5	12 (Low)	0 (0)	children	Surface water	Recreational park	NR	Not reported	Not reported	All clinical symptoms with or without microbiological confirmation
Study 2	20.9	31 (31.63)	16 (0)	Mixed	Food handler	catered meal	Sandwiches	Not reported	17 cases reported previous vaccination at anytime	All clinical symptoms
Study 3	14	24 (40)	7 (0)	Mixed	Food handler	Picnic	Potato Salad	Not reported	None	All clinical symptoms with or without microbiological confirmation

Study 4	17.4	7 (NR)	7 (0)	Mixed	Food handler	Restaurant	NR	Not reported	Not reported	All clinical symptoms
Study 5	21.4	24 (25.53)	24 (0)	Adults	Food handler	Outdoor activity	Cucumber salad	Not reported	All cases	Specific to fever with or without microbiological confirmation
Study 6	11.4	235 (34.97)	NR	Adults	Experimental	Experimental	Milk	10 ⁵	None	Specific to fever
Study 7	20.6	69 (65.09)	69 (0)	Mixed	Food handler	catered meal	NR	Not reported	Not reported	All clinical symptoms with or without microbiological confirmation
Study 8	17.7	26 (7.42)	26 (0)	NR	Unknown	catered meal	Pork meat	Not reported	Not reported	Specific to fever with or without microbiological confirmation
Study 9	14.7	35 (14)	35 (6)	Mixed	Food handler	catered meal	Vegetable	Not reported	Not reported	All clinical symptoms with or without

										microbiological confirmation
Study 10	10.5	84 (56)	77 (3)	Mixed	Food handler	catered meal	Spanish spaghetti	Not reported	Not reported	All clinical symptoms with or without microbiological confirmation
Study 11	9.7 & 7.6	24 (60)	0 (0)	Adults	Experimental	Experimental	Sodium bicarbonate	10 ³ & 10 ⁴	None	Specific to fever with or without microbiological confirmation
Study 12	9.1	41 (73.21)	NR	Mixed	Unknown	community	Corned beef	Not reported	Not reported	All clinical symptoms
Study 13	7	23 (79.31)	23 (3)	NR	Food handler	Picnic	Ice cream	Not reported	Not reported	Specific to fever and excluding diarrhea and vomiting

NR – Not reported

7.3.3 Test for heterogeneity

The p-value of the derived Friedman's statistic was <0.0001 and the calculated value of I^2 was 84% indicating high heterogeneity in the reported incubation period across the studies. The proportion of KS test p-values that were less than 0.05 was 63.6% (42/66) and the probability of obtaining the observed proportion was <0.0001 further confirming the high heterogeneity across the studies which is not due to chance.

7.3.4 Factors that may explain heterogeneity

Results of the univariate analysis showed that previous vaccination history, attack rate and year of study had a significant association with the mean incubation period. The mean incubation period reduced by 1.4 days with every 10% increase in attack rate (p-value 0.05), and outbreaks containing cases that had been vaccinated reported a longer incubation period of 9.8 days compared to outbreaks without vaccinated cases (Table 21). Outbreaks that occurred after 1950 reported a longer incubation period of 6.8 days compared to outbreaks that occurred before 1950. When these variables were included in the multivariable analysis, there was no significant association between vaccination, year of study and mean incubation period. The effect observed for year of study was reversed such that outbreaks that occurred after 1950 now reported a shorter incubation period of 3.9 days (Table 21).

Table 21 Linear regression analysis to identify factors associated with the distribution of incubation period in the review of incubation period of S. Typhi

Variables	Univariate analysis		Multivariable analysis	
	Difference in mean (95% CI)	p-value	Difference in mean (95% CI)	p-value
Age distribution				
Adult	Reference			
Children	2.7 (-8.2, 23.0)	0.6		
Mixed ages	1.5 (-6.6, 8.1)	0.6		
Year of study				
Pre 1950	Reference			
Post 1950	6.8 (2.4, 11.1)	0.01	-3.9 (-5.3, -2.5)	0.1
Attack rate	-0.14 (-0.25, -0.07)	0.05	-0.26 (-0.32, -0.20)	0.07
Setting				
Restaurant	Reference			
Catered meal	-0.5 (-20.7, 7.2)	0.9		
Community	-8.3 (-26.4, 4.1)	0.1		
Experimental	-7.4 (-26.6, 0.8)	0.2		
Outdoor activity	1.5 (-12.7, 18.7)	0.8		
Picnic	-6.9 (-24.7, 3.6)	0.2		
Food category				
Other	Reference			

Dairy	-4.3 (-13.1, 4.6)	0.3		
Non-food	3.1 (-6.9, 36.1)	0.6		
Red meat	0.1 (-9.1, 13.9)	0.9		
Salad & vegetable	3.3 (-5.5, 17.1)	0.5		
Vaccine history				
None	Reference			
Yes	9.8 (4.9, 17.7)	0.03	4.2 (2.2, 6.1)	0.1
Specific case definition				
No	Reference			
Yes	-2.2 (-7.9, 3.8)	0.4		

7.3.5 Identifying subgroups of studies

Based on the clustering analysis, studies were paired according to the evidence of dissimilarities between them. Studies found to have the least evidence of dissimilarities were paired and then connected by branches to another pair of studies or a single study with minimal dissimilarity to them. These pairing and connections of studies create a dendrogram of the dissimilarity matrix (Figure 26).

The pragmatic adjustments made to the significance level resulted in a corrected p-value of 0.004 and a corresponding cut-off point of 0.994. At this cut-off point, after considering multiple testing, four subgroups were identified. These included: one subgroup of three studies and three subgroups of two studies each (Figure 26).

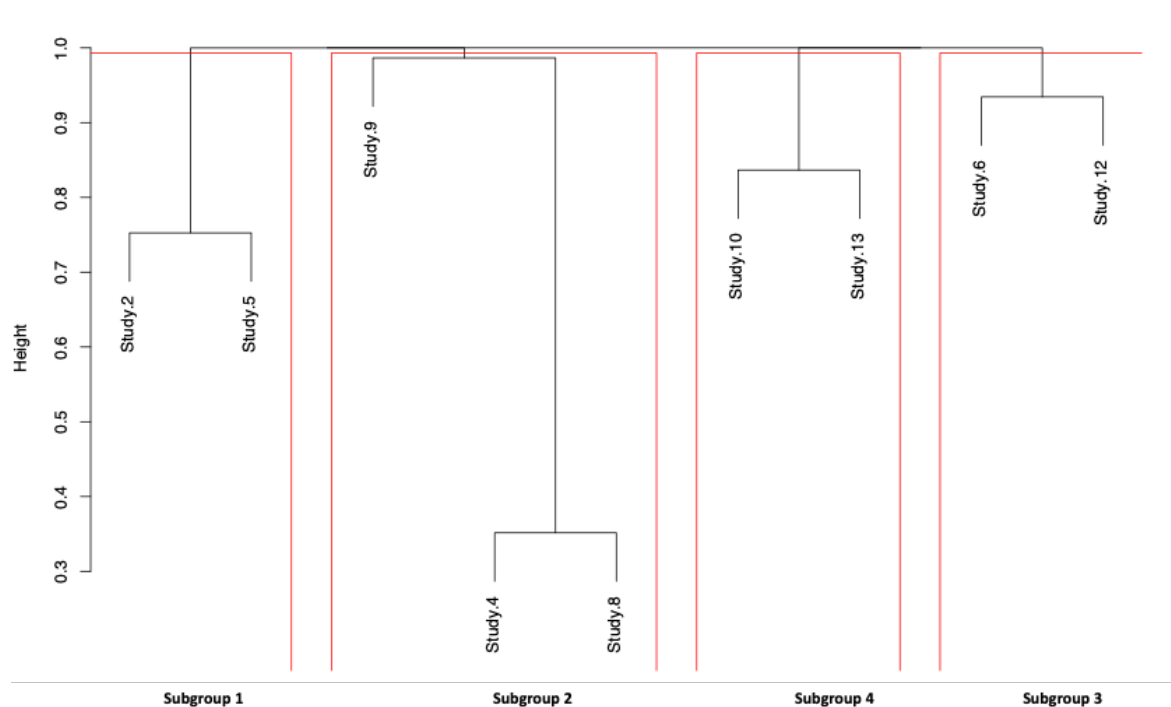


Figure 26 Dendrogram showing compact visualization of dissimilarity matrix and associated subgroups in the review of the incubation period of *S. Typhi*

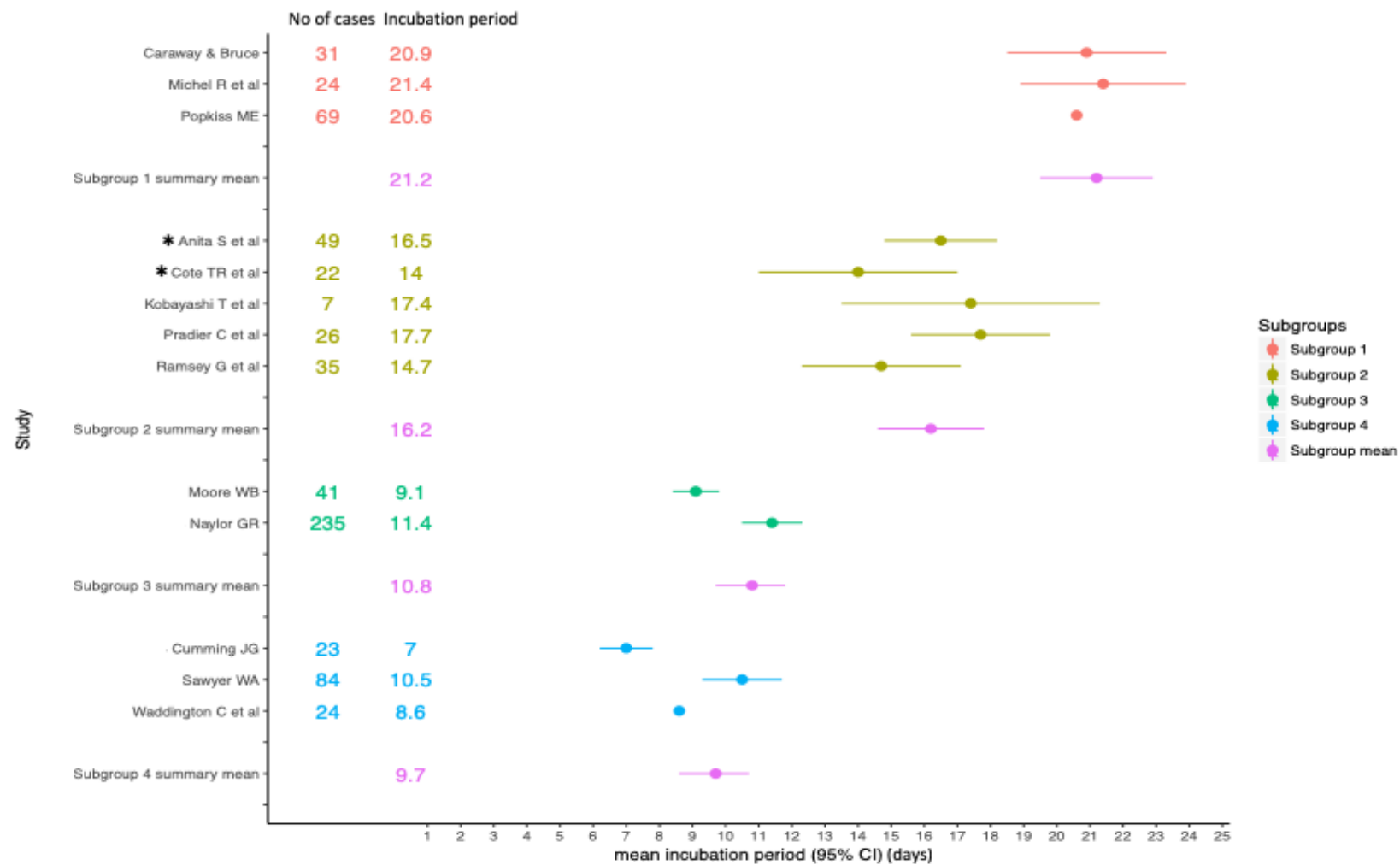
7.3.6 Summary of subgroup analysis

The mean incubation period was different between the subgroups (Figure 27) and significantly decreased from subgroup 1 to subgroup 3. Subgroup 4 had the shortest mean incubation period of 9.7 days (95% CI 8.6 – 10.7) and subgroup 1 had the longest mean of 21.2 days (95% CI 19.5 – 22.9) (Table 22). I also observed some differences in the variance, skew and kurtosis between subgroups. Subgroups 3 and 4 had the highest mean attack rate and also had the shortest incubation periods. Cases in subgroup 1 reported previous vaccination and also reported the longest incubation period.

7.3.7 Distribution of incubation period data

The Cullen and Frey graph (Appendix 12) indicated that the data were most suited to a beta distribution. However, this distribution is restricted to lower and upper limits respectively

which may not be biologically plausible. Further investigation of the use of such families of distributions will be necessary, but I limited this study to the gamma distribution which is more usually used to model incubation periods.



*Studies excluded from clustering analysis due to aggregated data

Figure 27 Forest plot showing mean incubation period and subgroup summary mean in the review of the incubation period of *S. Typhi*

Table 22 Summary statistics of outbreaks and incubation period within subgroups in the review of incubation period of *S. Typhi*

	Count	Sum of cases	Median	Mean (95% CI)	Variance	Skew	Kurtosis	Attack rate	Vaccine
Subgroup 1	2	55	20	21.2 (19.5 – 22.9)	42.2	-0.002	-0.6	28.7	All or most cases
Subgroup 2	3	68	16	16.2 (14.6 – 17.8)	42.7	0.9	2.3	Not reported	Unknown
Subgroup 3	2	276	9	10.8 (9.7 – 11.8)	39.7	2.1	4.4	54.1	None
Subgroup 4	2	107	8	9.7 (8.6 – 10.7)	29.3	1.8	2.9	67.6	Unknown

7.4 DISCUSSION

I have undertaken a systematic review to describe the distribution of incubation period of *Salmonella Typhi*. I tested for the presence of heterogeneity, attempted to identify factors that could influence the incubation period and classified the studies into subgroups with limited evidence of heterogeneity. Due to the high heterogeneity amongst the studies, estimating a single distribution of incubation period was impossible. However, I defined subgroups of similar studies and found their mean incubation period to range from 9.7 days to 21.2 days. These values are within the range quoted by international organisations reporting 3 to 60 days (63), 3 to 30 days (188) and also scientific literatures reporting 10 to 20 days (190). Some literatures report shorter incubation periods than I observed, like Raffatellu et al reporting a median of 5 to 9 days (191) or 8 to 14 days (190). I also identified attack rate and vaccination as possible factors that could affect the incubation period distribution.

The relationship between attack rate and incubation period is inversely proportional, such that higher attack rates result in shorter incubation periods (38,39). This has been reported in numerous outbreaks (129,156,193) and also observed in this review. Attack rates are associated with factors such as virulence of the organism, host characteristics and infecting dose (156). When either of these factors are present in a way that increases the attack rate: high virulence, susceptible host and large infecting dose, the incubation period is shortened as the onset of illness is quicker.

Vaccination is one of the preventive measures advised by public health professionals (194,195) to control or mitigate the burden of disease brought on by *S. Typhi* (64). It is recommended to people living in or visiting endemic areas in order to control outbreaks (63). This study did not aim to review the efficacy of typhoid vaccine in preventing outbreaks, however I identified a couple of studies with outbreaks reported in vaccinated population (196,197) and similar studies have also been reported (198). As observed from this review, it

may be stipulated that in populations where vaccination does not prevent illness, it may delay the onset of symptoms thereby prolonging the incubation period. This could have an adverse effect during outbreaks with continuous source as more people will be exposed and possibly infected before the outbreak is discovered.

The subgroup analysis indicated that outbreaks and experimental studies that took place before the year 1950 belonged to subgroups reporting shorter incubation period (subgroups 3 and 4). In the univariate analysis, post-1950 studies had a significantly longer incubation period, however, in the multivariate analysis, this effect on the incubation period was reversed as we observed a shorter incubation period in post-1950 studies. The change in the direction of effect is indicative of the presence of a confounding factor in the univariate analysis.

A lot of reported outbreaks caused by *S. Typhi* either had a continuous source of exposure such as municipal water supply (199–202), or were caused by widely distributed food produce (203,204). This means that point source outbreaks, where the time of exposure is definite and incubation period could be accurately calculated, were few. This was a similar finding in the review conducted by Naylor (193). The author stated the criteria of an outbreak suitable for incubation period analysis and reported that very few reports fulfilled these criteria. However, in their review, studies where the date of exposure was not confirmed were included and date of infection was inferred from date of purchase, date of sale for a batch of food item, date of water contamination and so on. Continuous source outbreaks can be useful if the investigators identify and report the exact dates of exposure and onset but this was not the case in the outbreaks reviewed by Naylor (193). Hence, some studies identified in Naylor's review were not included in this review. In this study, I identified only thirteen studies with point source exposures or where the time of exposure was known. Some studies however reported point source exposures, although the exposures could have occurred over a few days (205,206).

Dose response is known to be associated with incubation period (192,207), and although this was not particularly studied in this review, there were studies where this relationship was observed. In an outbreak associated with ice cream (189), a dose response relationship was observed as those who had a mixture of ice creams had longer incubation periods while those who had only the contaminated ice cream had comparatively shorter incubation periods. Another dose response relationship was observed in the experimental study conducted by Waddington et al (65). Two cohorts of volunteers were infected with different doses of inoculum, 10^3 and 10^4 . The cohort receiving the lower dose had a longer incubation period compared to those receiving the higher dose.

The case definition used in identifying cases or in including cases as part of an outbreak is very subjective and based on what the authors identify as a case. This is quite essential because it can either prolong or shorten the incubation period of a case based on what is considered as the onset of symptoms. Although there are guidelines on what to consider when defining a case (6), case definitions depend on the authors' perception. The complex syndrome of typhoid fever (185) makes the process of defining cases arbitrary. In this review, some of the case definitions were broad and syndromic, such as 'clinical illness', while others were very specific in terms of fever at a particular temperature (65,197,208) or dismissing symptoms such as diarrhoea and vomiting (189). Neither the broad nor specific definitions are incorrect, however, the varied case definitions may have contributed to the observed heterogeneity.

A limitation I encountered in this review was the difference in the reported incubation period data extracted from the included studies. Although the raw incubation times were reported in days, some studies reported an aggregated date of symptom onset in three or seven-day intervals (209,210). Onset times were either rounded up or rounded down to fall into the reported interval. Rounding down the onset time from three days to one will spuriously shorten the incubation period just as rounding up the onset time from one day to seven will spuriously prolong the incubation period. I therefore excluded these studies from the

hierarchical cluster analysis and their reported mean incubation periods should be interpreted with caution.

7.5 CONCLUSIONS

This study showed that the reported incubation period varied more than could be explained by chance and identified attack rate and previous vaccination as associated factors, although not significantly. This may be due to the limited information available from the small number of studies reviewed. Analysing individual patient data will provide an opportunity to assess additional patient characteristics such as underlying medical conditions, drug interactions and full vaccination history. In addition, using mathematical models to describe the process of infection will help identify parameters intrinsic to the infection pathway that could influence the distribution of incubation period such as gastric transit times, phagocyte characteristics and cellular bacteria growth.

In order to maximise the use of outbreak reports in studying parameters such as incubation period, authors should endeavour to precisely report exposure times and onset times. Where the outbreak is non-point source, identifying and reporting the exposure time of each case where possible will be useful. Case definitions should be developed in such a way as to ensure that cases are included based on similar thresholds which can also be comparable with other outbreak reports.

CHAPTER 8: REVIEW: PATHOPHYSIOLOGY OF *SALMONELLA* TYPHI AS IT RELATES TO THE DISTRIBUTION OF INCUBATION PERIOD.

ABSTRACT

Typhoid fever is a systemic infection caused by the bacteria *Salmonella* Typhi. It has a longer incubation period and is characterised by different clinical syndrome compared with non-typhoidal salmonellosis. Available reports on the incubation period of typhoid fever are inconsistent, and one of the factors contributing to this inconsistency is the individual variation in host-pathogen interaction. Understanding the relationship between the host and the pathogen, and how this contributes to the distribution of the incubation period will support outbreak investigations and contribute to the reduction of the burden of disease.

I conducted a literature review to summarise the pathophysiology of *S. Typhi* with reference to the duration of each step in the process of infection. Available evidence on duration and rates were collated to inform parameters. Based on a developed hierarchy of evidence and reliability, I prioritised and selected parameter values that can be used for research purposes.

The pathophysiology of *S. Typhi* was summarised into thirteen steps describing the process of infection and estimating the duration of each step from ingestion to secondary bacteraemia.

8.1 INTRODUCTION

Salmonella enterica serovar Typhi (S. Typhi) are bacteria that cause the systemic infection typhoid fever, usually following ingestion of contaminated food or water (191).

Gastroenteritis is not a typical symptom of S. Typhi as infection goes beyond invasion of the gastrointestinal system and affects the systemic organs (191), particularly the spleen and liver resulting in splenomegaly and hepatomegaly (185). Reported symptoms are usually systemic such as fever (182), headache, loss of appetite and malaise. Infection can be quite severe if left untreated and could result in case fatality rates of about 1% (19).

The incubation period of typhoid fever, the time between exposure to the pathogen and the onset of clinical illness, varies among individuals as a result of host and pathogen characteristics and perhaps factors particular to each infecting event such as the food matrix. Incubation period distributions of 6 – 30 days (211), 12 – 16 days and 13 – 22 days (66) have been reported. The systematic review conducted to describe the incubation period of S. Typhi reported a mean incubation period between 9.7 days and 21.2 days within subgroups (Chapter 7). This incubation period is substantially longer than that of most other serotypes of *Salmonella* where it is typically reported to be between 6 – 72 hours (212,213).

Knowledge of the pathophysiology of S. Typhi contributes to understanding the incubation period distribution, likewise, accurate knowledge of the incubation period improves our understanding of the pathogenesis. Based on this ongoing knowledge cycle, getting insight into the host and pathogen interaction is necessary.

In addition to understanding the pathophysiology and dynamics of the pathogen, accurate knowledge of the incubation period informs outbreak investigations by providing necessary intelligence and contributes to the reduction of the burden of disease of S. Typhi. Accurate estimates of the distribution of incubation period are also necessary to distinguish secondary cases from primary cases, formulate a hypothesis, and determine the end of an outbreak during outbreak investigations (55).

In this chapter, I conducted a literature review and summarised the pathophysiology of *S. Typhi* with reference to the duration of each step in the process of infection. I collated all available published literature on the duration of each step in the pathophysiology of Typhoid fever. For each step, data to inform parameter estimates were provided and the evidence around the reported data was presented. Where there were several studies reporting different estimates for the same step, the most reliable estimate was selected using the approach described in methods. Where possible, I made use of human experimental studies but supplemented with mouse models where human studies were not available.

8.2 METHODS

8.2.1 Literature search

Literature searches were conducted to achieve the following objectives:

- to identify studies reporting the pathophysiology of *S. Typhi* in order to describe and summarise the steps involved in the infection process;
- to identify reports on the different steps of infection in order to estimate the duration of each step.

In identifying studies reporting the pathophysiology of *S. Typhi*, related keywords including 'typhoid' or '*Salmonella Typhi*' and 'pathophysiology' were used. Studies reporting the steps of interest were identified using specific keywords. For instance, to identify studies reporting gastric emptying time, the keywords used included 'gastric' or 'stomach' and 'emptying'. An exploratory search was conducted in Google and Google Scholar using limited keywords and reviewing the first two pages for relevant articles. Where there was little or no information on the supporting evidence for a step, the search was escalated to include book chapters, review papers and references of already identified articles. All searches were conducted between January and September 2017.

8.2.2 Summarising the pathophysiology and extracting parameter values

I summarised the pathophysiology of *S. Typhi* from studies identified in the literature search and broke down the process of infection into steps based on the location of the bacteria in the body. For each step of the infection process, the key mechanisms were described and parameter values were estimated representing the duration of infection or rate of infection. Each study was reviewed to identify the type and number of experiments conducted, the number and characteristics of study participants, the unit of reported values (duration, rate or proportion), and the distribution of the parameter values. The measure of central tendency (mean or median), variation (standard deviation or range), where available, the raw data and any information on the shape of the distribution were extracted.

8.2.3 Hierarchy of available evidence for parameter values

The reliability of a parameter value depends on the quality of the supporting evidence available. The highest level or the best quality evidence of parameter values for incubation period will be identified from large in-vivo human experiments using *S. Typhi*. As this is not usually possible, other human and animal experimental studies, mathematical model simulations and biological estimations are used to derive parameter values with varying levels of reliability based on the quality of evidence. Where necessary, in-vivo and in-vitro mouse studies of *S. Typhimurium* were used to describe cellular infection as replication and death of bacteria and their interactions with macrophages are similar in mice and humans. Figure 28 shows a schematic representation of the types of evidence available and how they are ordered as a hierarchy to inform which evidence I prioritised.

8.2.4 Selecting parameter values

Where several studies were identified reporting different values for a particular step in the infection process, I chose an estimate as well as reported other data. The selection was based on the reliability of the supporting evidence according to the recommended hierarchy, the characteristics of the study conducted in terms of study participants and units of measurement. In addition, the practicality of the estimated parameter value in developing mathematical models was also considered when making suggestions. To this end, I attempted, where possible, to change the estimate to a common time unit and measure to aid the incorporation of the estimated parameter values into mathematical models.

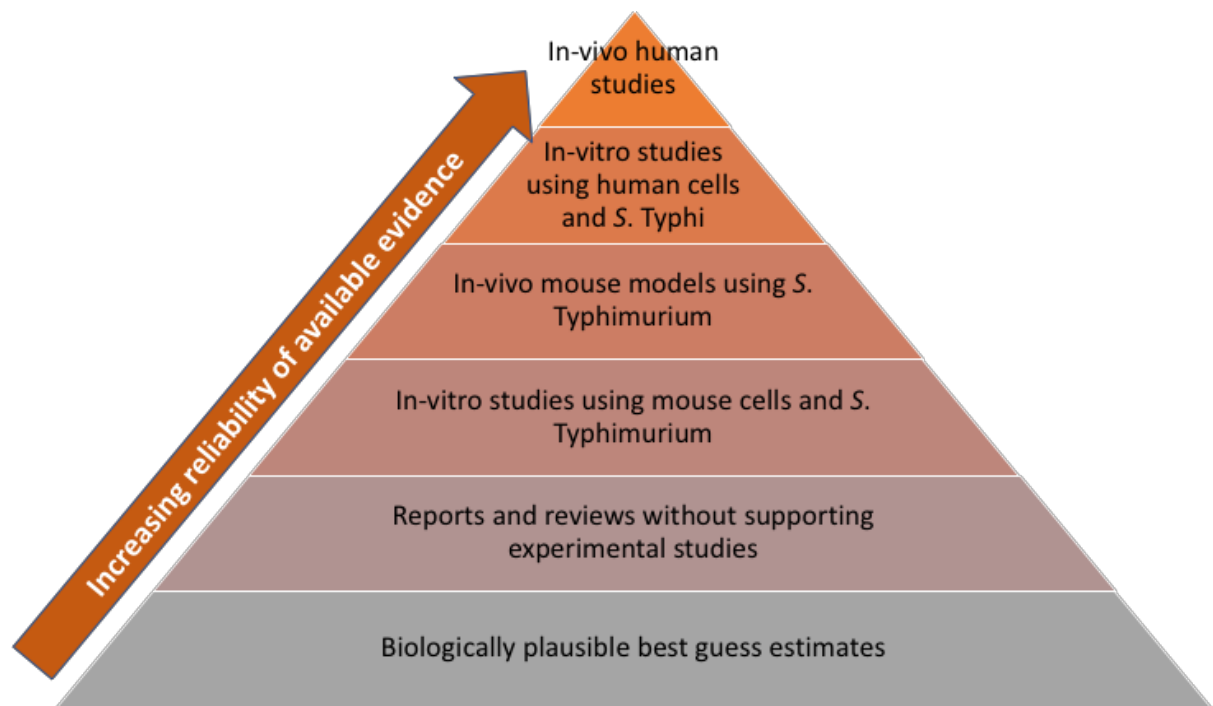


*Figure 28 Hierarchy of evidence for parameter values derived from the literature in the review of the pathophysiology of *S. Typhi**

8.3 RESULTS

8.3.1 Pathophysiology of *Salmonella* Typhi infection

During the process of *S. Typhi* infection, bacteria travel from the mouth through a range of body sites where they either transit through, replicate within or exit the infection pathway through death or excretion. The identified steps in the infection process are summarised in Figure 29. The detectable presence of bacteria in the blood coincides with the onset of clinical illness (214,215), and the combined duration of each step results in an estimation of the incubation period.

8.3.2 Duration of the identified steps in the process of infection

A total of thirteen steps indicating the process of infection were identified from the literature. A brief description of each step is given, evidence on duration and rates are presented and rationale supporting each evidence selection is specified.

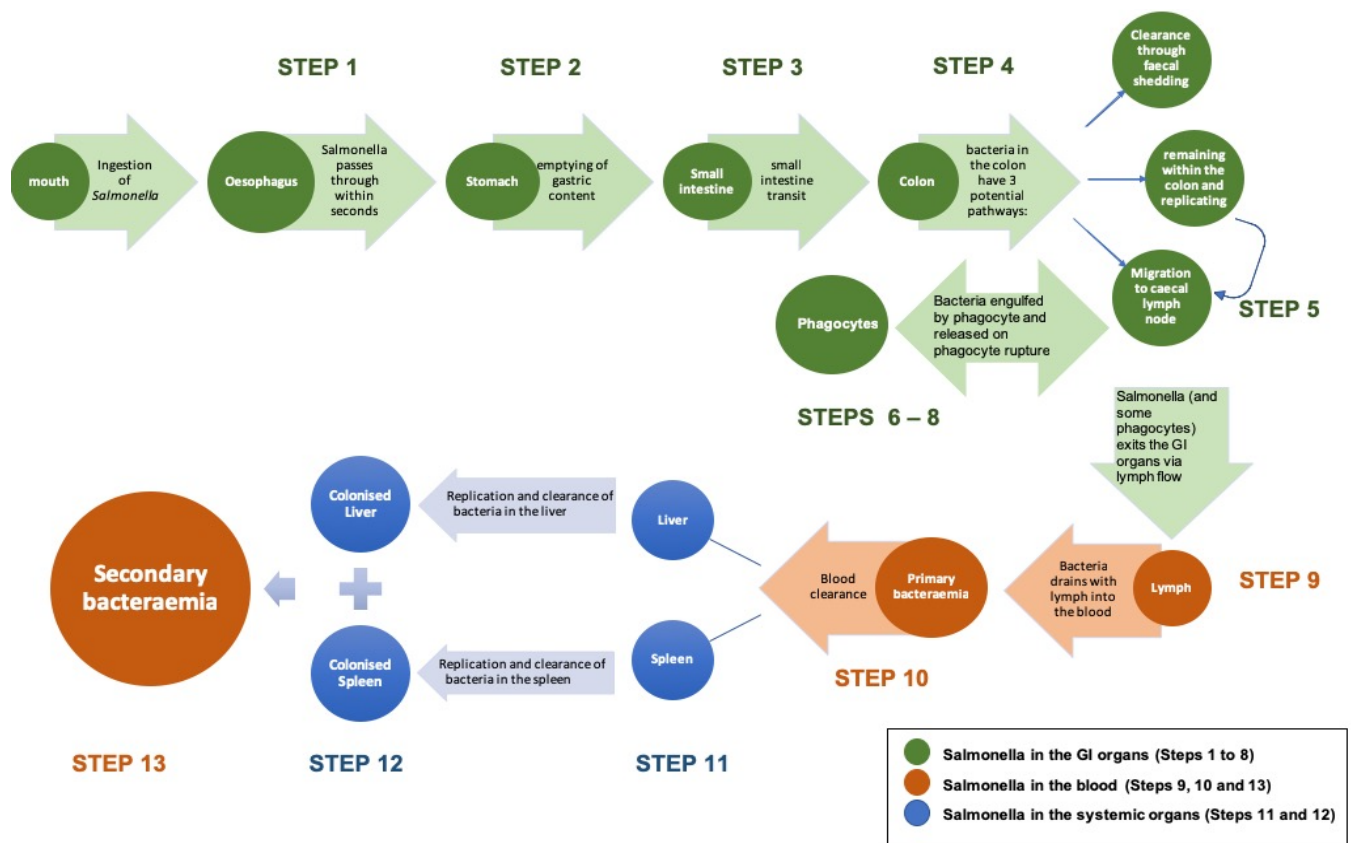


Figure 29 Theoretical flowchart showing steps of the infection process of *Salmonella* Typhi

Step 1 – Ingestion of bacteria

This is the point of exposure to the infecting pathogen and the beginning of the infection process as well as the start point of the incubation period. When food is swallowed, it travels through the oesophagus and arrives at the stomach within a few seconds. The oesophageal transit time for solid and semisolid food is estimated to range between four seconds to eight seconds, while liquid transits in approximately one to two seconds (216). A liquid meal of 15 to 30 mL will transit through the oesophagus in about 6 – 15 seconds (217).

Step 2 – Gastric emptying

After the bacteria is ingested, they travel to the stomach where the resident microbes and stomach acid eliminate some bacteria. The remainder of the bacteria along with the gastric contents empty into the small intestine. The gastric emptying time varies depending on the consistency of food (liquids or solids) and the composition of food (ratio of carbohydrate, protein and fat) that is ingested (218). The emptying time for solids is significantly slower. Gender has also been observed to affect gastric emptying rate of solids (219). In a human experimental study conducted by Bennink et al (220), the mean $\pm$ SD of 10% liquid emptying time, which is the time it takes for 10% of liquid gastric content to be emptied, for males and females was reported as 6.4 ± 3.3 min and 3.9 ± 1.8 min respectively, while the 10% solid emptying time was 14.3 ± 7.1 min for males and 20.7 ± 8.1 min for females. The 50% liquid emptying time (half time) for males and females was 35.4 ± 12 min and 29.8 ± 9.6 min respectively, while the solid emptying half time for males and females were 50.5 ± 13.6 min and 84.0 ± 24.7 min respectively. The overall mean gastric emptying rate of liquid content was reported as 2.3 ± 1.0 %/min for males and 2.3 ± 1.2 %/min for females while the mean gastric emptying rate for solid content was 2.3 ± 0.6 %/min for males and 1.4 ± 0.6 %/min for females. This experimental study compared the gastric emptying time between solids and liquids in 24 volunteers (12 males and 12 females) using static imaging (220). Images were taken every 10 mins for the first one hour and every 15 mins for the next seven hours.

In another human experimental study (221), the mean $\pm$ SD emptying half time for liquids in males and females was reported as 81.6 ± 21.5 min and 79.4 ± 22.9 min respectively. The 50% emptying time for solids was reported as a median of 122 min (25 – 75% percentile: 111 – 182 min) in males and 131 min (113 – 167 min) in females. The study used ^{13}C breath tests to measure gastric emptying times for liquids and solids in 90 volunteers (45 males and 45 females). Breath samples were collected every 10 min for the first 120 min and then at 135, 150, 165 and 180 min after ingestion. No correlation was found between gastric emptying times and age, gender and body mass index.

In a multicentre study of 189 volunteers (99 females and 90 males) (219), the median gastric emptying half time for solid content was 63.1 min (5 – 95th percentile: 41.0 – 96.3 min) for males and 74 min (48.2 – 115.4 min) for females. This study used scintigraphy to measure the emptying time. For the first 30 mins after the meal, continuous imaging was taken, and then at 30 min or one-hour interval for the next four hours.

The difference in transit rate between solids and liquids and between gender, which is consistent across all three experiments, may result in an inconsequential difference in incubation period, based on the implicating food vehicle or the exposed population. For research purposes, the values used will depend on the characteristics of the research participants, however, I have selected the estimates reported by Bennink et al (220). This was the only study that reported the emptying rate for 100% of gastric content as well as the 10% and 50% emptying rates. Although the study population was small, data was collected frequently over eight hours, resulting in over 800 observations. This detailed reporting would make these values useful in modelling work as the value of the parameter can be changed accordingly.

Step 3 – Transiting through the small intestine

After the gastric contents empty into the small intestine, the bacteria travel through to the distal ileum and caecum. The small intestine transit time does not vary by food type, and gender does not affect the transit time (218,220,222,223), however, some individual variability is expected in the observed transit time. Small intestine transit time is calculated as the difference between the time 10% of gastric contents entered into the duodenum and the time it reached the caecum.

In a study conducted by Degen & Phillips (223), 32 healthy volunteers (12 women, 20 men) were monitored in two experiments to identify the variability of gastrointestinal transit times. The mean small intestine transit time in females was reported to be 196 ± 22 min in study 1 and 210 ± 23 min in study 2. In males, the transit time was reported to be 181 ± 19 min in study 1 and 196 ± 13 min in study 2. A total of nine images was obtained per study per volunteer over a 48-hour period.

Using scintigraphic method to assess the transit time in 19 volunteers (10 males and 9 females) (224), the reported mean transit time was 220.9 ± 49 min with a range of 131 to 322 min. In another study using scintigraphy to assess transit times in 30 healthy adults (14 males and 16 females) (225), the mean transit rate was reported as 77.9 ± 31.1 min with a range of 21 to 153 min.

Yu et al conducted a study to parameterise compartment and dispersion models in order to describe the small intestine transit process. Using the descriptive analysis of over 400 human transit time data, they estimated the small intestine transit time. The compartmental model predicted the number of compartments in the small intestine and estimated the transit time through each compartment. The dispersion model estimated the velocity and dispersion coefficient of the small intestine. The small intestine total transit time was reported to be a mean of $199 \text{ min} \pm 78 \text{ mins}$ and a median of 191 min (range 30 - 570 min). The transit times in the duodenum, jejunum and ileum were 14, 71 and 114 min respectively. The descriptive

analysis indicated a wide variation in the reported values of the small intestine transit rate which may indicate substantial individual variations in this parameter.

The studies reviewed reported similar mean transit times of approximately 200 min, however, I selected the values reported by Yu et al (226) as this involved a large sample size enabling a more accurate measure of the spread of data (variance). The compartmentalised transit times as well as total transit time were also estimated. The study was conducted to parameterise models, making the output suitable for modelling work. Furthermore, the actual transit times of each study participant can be deduced from their study making additional analysis of the data possible.

Step 4 – *Salmonella* in the colon (replication and shedding)

Almost all the bacteria travel to the ileo-caecal junction, without encountering any detectable restrictions (227), where they can either replicate, be expelled through faeces or migrate to the caecal lymph node (CLn).

At the ileo-caecal junction, bacteria invade the epithelial cells and M-cells in the Peyer patches (182) with the aim of migrating to the CLn (214). Within M-cells in the Peyer patches, some bacteria are killed by resident macrophages and dendritic cells in the submucosa, while the remaining replicate.

The average doubling time for the bacterial population is approximately 95 min, however a sub-population, referred to as hyper-replicating *Salmonella*, doubles every 20 min (228,229). This was reported in an in-vitro experiment where cultured human epithelial cells were infected with *S. Typhimurium* (229). This ongoing replication increases the bacterial population size in the gut compared with the ingested dose (227).

S. Typhi is invasive but does not usually trigger a rapid inflammatory or diarrhoeal response (182,230), however, it is frequently shed in the stools of infected persons (192), thus, some

of the bacterial population is lost. In sub-clinical stages of typhoid, shedding occurs intermittently and becomes more persistent in untreated cases (182). The quantitative microbiology of *Salmonella* in stool during acute infection is relatively unknown and I was unable to identify any evidence in the literature. No inference could be made on the rate and quantity of bacterial shedding. Until this evidence gap is filled, it is necessary to apply biologically plausible values.

Step 5 – Migrating to the caecal lymph node

Salmonella Typhi possesses virulence traits that enable it overcome mucosal barrier functions and spread from the gastrointestinal tract to the lymph nodes (191,231), thus, a fraction of the bacteria migrate to the caecal lymph node (CLn). Invasion of the CLn, indicating the beginning of systemic infection, occurs through stochastic immigration, replication and clearance of bacteria.

According to Kaiser et al (227), each bacteria cell has a chance of 3×10^{-7} (95 % CI: 2.4×10^{-7} to 3.9×10^{-7}) to migrate to the CLn 24 hours after infection. This value was estimated from an in-vivo mouse experimental study using wild type isogenic tagged strain (WITS) of *S. Typhimurium* to study the migration rate and growth rate of bacteria in the CLn. Mice were orally inoculated with a dose of 5×10^7 CFU, which then increased to a population of 10^9 in the colon (227). Seventy-seven observations were derived from using seven types of WITS in eleven mice. A mathematical model with a stochastic birth-death process extended by immigration was then employed to analyse the infection data and determine the rate of migration and distinguishing between the contributions of migration and replication in the final bacterial population of the CLn. This very low migration rate creates a major bottle neck for infection, considering the bacterial population in the colon.

Step 6 – Invasion of phagocytes

The process of invading the CLn is a complicated process involving resident phagocytes through a process known as phagocytosis. The bacterial population is split between intracellular (within the phagocytes) and extracellular (outside the phagocytes but within the CLn) environments. When bacterium encounter a phagocyte, it attaches for a period of less than ten seconds, and longer attachments indicate a partial engulfment. Bacteria that attach to the cell for longer than ten seconds have a high chance of entering the cell (232).

In order to examine the process of phagocyte infection, Gog et al used both a physical model and a mathematical model to estimate the infection rate of a phagocyte (232). In vitro experiments were carried out using phagocytes derived from the bone marrow of mice and infecting them with *S. Typhimurium* at different multiplicity of infection (MOI – ratio of bacteria to phagocyte) between 10 and 800. The experiments were conducted in triplicates resulting in a total of 9,000 observations.

According to the physical model, the rate of phagocyte invasion was 2.5×10^{-5} per second per bacterium averaged across MOI. From the mathematical model, the estimated invasion rate was 4.2×10^{-5} per second per cell per MOI (232). These rates differ mainly due to the broad approximations made in the physical model and the difficulties in parameterisation in the mathematical model. Determining the actual MOI during an infection is extremely difficult, therefore, I selected the rate derived from the physical model for better estimations.

Step 7 – Growth of bacteria in the phagocytes

After invasion, bacteria may be killed by resident phagocytes in an attempt to clear the infection or they may replicate and induce cell death in phagocytes, causing them to rupture (231,233). Its ability to persist in the host cell is crucial for the process of infection to continue.

In the phagocyte, the engulfed bacteria are encased in a membrane composed of host cell

called a *Salmonella* containing vacuole (SCV) (234,235) where they reside and proliferate. Bacteria begin to replicate intracellularly two hours after phagocyte invasion and would continue to replicate for up to three days if the phagocyte survived that long (236).

In an experiment conducted by Lowrie et al (237), the doubling time of bacteria was 2.9 ± 0.7 hours when measured by light microscopy and 2.3 ± 0.4 hours when measured by CFU counts. In this in-vitro experiment, murine macrophages were infected with *S. Typhimurium* and penicillin-induced death rate was used to estimate the bacterial division rate.

Forest et al conducted an in-vitro experiment to observe intracellular growth of bacteria using human cells (236). Human cultured macrophages were infected with *S. Typhi* and the number of intracellular bacteria was determined from 2 to 96 hours after infection. Three independent experiments were performed in duplicates and the mean number of bacteria was reported at each time interval. At six hours post infection, the number of bacteria had increased by 140% and by 24 hours post infection, there was a 200% increase in the number of bacteria. As this is an experimental study using human cells, output from this study is selected. Data on proportional increase of bacteria can be extracted and fitted to a replication model to estimate the growth rate.

Step 8 – Phagocyte death and rupture

The mechanism by which bacteria enter the phagocyte triggers the process of *Salmonella* induced programmed cell death (238). Normally, the presence of a foreign body within a phagocyte activates host cell immune response in an effort kill the bacteria, however, using a unique secretion system, type III secretion system, *Salmonella* injects proteins that alter the cell structure and halts the cell's immune response (235) and also causes cell death (239). Within 40 minutes of phagocyte invasion, infected cells begin to shrink and develop SCV and the process of cell death starts (233). Three hours after infection, the cell membrane ruptures or collapses, releasing the intracellular bacteria (240).

An in-vitro experiment was conducted by Monack et al to examine the cytotoxicity of *S. Typhimurium* on murine phagocytes (238). Phagocytes were infected at a MOI of 10 and the proportion of dead phagocytes were counted at 30, 60, 90 and 120 minutes after infection. At 60 minutes post infection, an average of eight percent of infected phagocytes were dead. The proportion of dead phagocytes increased at 90 min and 120 min post infection to an average of 18.4% and 26.3% respectively. Data from this study can be extrapolated to determine how long it will take for 50% or 100% of phagocyte cells to rupture. When the phagocytes rupture, the resident bacteria re-enter the free environment in the CLn to invade other phagocytes.

Step 9 – Exiting the GI organs

The caecal lymph node serves as a portal for systemic dissemination of bacteria. *Salmonella* leaves the GI organs through lymph that empties into the blood (241) and causes an infection that is clinically (182) and frequently microbiologically undetectable (214). There is no evidence on the proportion of bacteria that exit the lymph node per unit time, so we have assumed that there is continuous flow of bacteria through the lymph as long as the CLn is infected.

We can assume that bacteria will drain from the caecal lymph node at the same rate that lymph flows through the intestinal lymphatic vessel. There is little quantitative information on the lymphatic system available in the literature. A report on the lymphatic system estimated that for an adult male weighing 70kg, the intestinal lymphatic vessel has a flow velocity of 3400µm/sec and a corresponding flow rate of 14mm³/sec (242).

In a review of the lymphatic structure and function in the different GI organs, the lymph flow in the large intestine was reported to be approximately 0.015ml/min/100g (243).

Step 10 – Transit to systemic organs through blood

There is no death and no replication of bacteria in this step, only transit to systemic organs

and a transient primary bacteraemia (77). The presence of bacteria in the blood results in additional tissue, particularly the spleen and liver, being colonised. We can assume that bacteria will travel into these organs at the same rate of blood supply to the organs.

The liver receives about 70% of its blood supply via the portal vein, which is also the direct venous outflow from the intestine (244). In a human experimental study, the portal venous blood flow was measured in 45 volunteers (26 males and 19 females) using doppler sound. Measurements were taken while the volunteers were placed in three different positions: standing, in a supine position and lying with head tilted down. The mean blood velocity $\pm$ SD in the portal vein while standing was 13.9 ± 4.49 cm/s, while in supine position was 12.3 ± 5.9 cm/s and with the head tilted was 11.8 ± 2.9 cm/s. The corresponding mean volume flow for the different positions were 662 ± 169 ml/min, 864 ± 188 ml/min and 831 ± 170 ml/min respectively (245).

Perišić et al (246) measured the splenic velocity in 25 healthy volunteers using a doppler system and reported a mean systolic splenic artery velocity of 58.5 ± 13.3 cm/s and a diastolic velocity of 19.7 ± 5.5 cm/s. The mean blood flow of the splenic artery was not reported making data from this study less suitable.

Another study conducted by Sato et al measured splenic artery blood flow using a doppler system and a more invasive electromagnetic blood flowmeter in 24 healthy subjects (247). The mean blood velocity $\pm$ SD was measured to be 18.7 ± 4.2 cm/s, and the corresponding mean blood flow was 179 ± 37 ml/min or 3.3 ± 0.7 ml/min/kg. Besides from reporting the mean blood flow, I selected this study because two methods of testing were used making the results more reliable.

Step 11 – Infection of systemic organs

The systemic spread of *Salmonella* has been extensively studied in an in-vivo mouse experiment using wild-type isogenic-tagged strains (WITS) (77). Bacteria strains were given

a different DNA signature tag that enabled them to be monitored at different sites in the body of mice during systemic infection. The mice were injected intravenously to mirror the onset of primary bacteraemia and the growth kinetics of the WITS in the systemic organs were monitored. Total bacterial load in both organs were measured at different time intervals. Thirty minutes after the onset of primary bacteraemia, bacteria are observed in the spleen and liver (231) where they infect the phagocyte one bacterium per cell (77). Due to phagocytic action, the number of viable bacteria reduces (231) decreasing by 50% every five hours in the spleen and every three hours in the liver (77). As there was minimal migration between organs, this decrease was mainly due to bacterial death (77,231).

Step 12 – Colonisation of systemic organs

Occurring simultaneously with bacterial death, is replication within phagocytes which plays a major role in systemic organ colonisation (248). Similar to the process in the CLn, bacterial population increases by replicating within phagocytes, lysis of infected phagocytes and release of bacteria to extracellular compartment to infect other resting phagocytes (77,231). Within the first six hours of infection, high bacterial replication and an even higher death was observed resulting in a slow decline in bacterial population in the systemic organs. During this time, the bacterial population doubled every 1.7 hours and had a half-life of 1.1 hours. This phase of high death rate ceased within seven hours and was followed by a phase of moderate bacterial growth and negligible death in the spleen and liver which occurred between 6 to 24 hours after primary bacteraemia (77). The bacterial population doubled every eight hours and the bacterial death was negligible. The host defence mechanism at this stage was more bacteriostatic, to restrain the growth of bacteria, rather than bactericidal. Intracellular bacterial numbers were skewed towards low numbers with a greater proportion of phagocytes containing very few bacteria.

Step 13 – Re-entry into the blood and onset of secondary bacteraemia

After a period of replication, bacteria re-enters into the blood stream. According to Grant et al (77), *Salmonella* exits the systemic organs at a fraction (1%) of the initial entry rate (described in step 10) and returns into the blood stream. This typically occurs from about 48 hours post primary bacteraemia, and bacteria mortality remains negligible.

This final step in the infection process indicates the end of the incubation period (214) and the onset of clinical symptoms and is characterised by a now detectable secondary bacteraemia and further systemic bacterial spread.

Table 23 Summary table of collated parameter estimates and type of evidence available for each step in the infection process

Step in infection process	Estimated parameter value (cited reference)	Range	Type of available evidence	Reference for other available literature
Step 1	< 15 sec (216,217)	0 - 15	Biologically plausible estimate	
Step 2 - liquid in males	2.3 %/min (220)	SD $\pm$ 1	Human in-vivo studies	(219,221)
Step 2 - liquid in females	2.3 %/min (220)	SD $\pm$ 1.2	Human in-vivo studies	
Step 2 - solids in males	2.3 %/min (220)	SD $\pm$ 0.6	Human in-vivo studies	
Step 2 - solids in females	1.4 %/min (220)	SD $\pm$ 0.6	Human in-vivo studies	
Step 3	3.31 hours (226)	SD $\pm$ 78	Human in-vivo studies	(223–225)
Step 4 - replication in colon	1.5 hours (229)		Human in-vitro study with <i>S. Typhimurium</i>	
Step 4 - shedding from colon	No evidence available		No evidence available	
Step 5	3 X 10 ⁻⁷ /day (227)	95% CI - 2.4 X 10 ⁻⁷ to 3.9 X 10 ⁻⁷	In-vivo mouse study using <i>S. Typhimurium</i>	
Step 6	2 X 10 ⁻⁵ /sec /bacterium (232)		In-vitro mouse study using <i>S. Typhimurium</i>	
Step 7	Data to be fitted to model (236)		In-vitro human study using <i>S. Typhi</i>	(237)
Step 8	Data to be fitted to model (236)		In-vitro mouse study using <i>S. Typhimurium</i>	(237)
Step 9	840 mm ³ /min (242)		Report and reviews without supporting evidence	(243)
Step 10 - Portal vein	662 ml/min (245)	SD $\pm$ 169	Human in-vivo studies	
Step 10 - Splenic artery	179 ml/min (247)	SD $\pm$ 37	Human in-vivo studies	(246)
Step 11 - Doubling time	1.7 hours (77)		In-vivo mouse study using <i>S. Typhimurium</i>	
Step 11 - Half life	1.1 hours (77)		In-vivo mouse study using <i>S. Typhimurium</i>	
Step 12 - Doubling time	8 hours (77)		In-vivo mouse study using <i>S. Typhimurium</i>	
Step 12 - Half life	Negligible (77)		In-vivo mouse study using <i>S. Typhimurium</i>	
Step 13	1% of value in step 10 (77)		In-vivo mouse study using <i>S. Typhimurium</i>	

8.4 DISCUSSION

Knowledge of the pathophysiology of *S. Typhi* and the underlying distribution of the parameter values, as collated in this review, provides an opportunity to understand the process of infection and how each step contributes to the distribution of the incubation period of *S. Typhi*. It can also support the development of a within-host mathematical model to understand the incubation period of the pathogen and identify host and pathogen factors that can influence the distribution. It will form the basis of a compartmental model that simulates the infection process from exposure to onset of illness and also inform model parameterisation as rates and proportions can be deduced. Although this review was particularly tailored to inform incubation period, it can also be used to inform dose response models on incubation period.

The accuracy or reliability of any model output depends on the precision of the parameter values, and the strength of evidence on which the parameter value is based depends on its source. Parameter values are mostly derived from the literature and possible sources include in-vivo human and animal studies, in-vitro human and animal studies and reviews. In some cases, there is limited evidence available in the literature, hence, a best guess estimate, which is also biologically plausible, is made.

The parameter values identified in this review were inferred from numerous sources, with varying strength of evidence (Table 23). A number of parameter values were derived from in-vivo and in-vitro studies of human and animals, using either *S. Typhi* or *S. Typhimurium*, thereby having strong supporting evidence, albeit some stronger than others, which should result in reliable model outputs. However, parameter values identified in steps 4 and 9, relating to the shedding rate of *S. Typhi* and the flow rate of lymph respectively, had little or no evidence available in the literature. Although there was evidence on the frequency of stool shedding during infection (182), I was unable to identify any reports on the quantitative microbiology of *S. Typhi* in faeces. For the purpose of research, a best guess estimate of the shedding rate, based on the authors' understanding of what happens to *Salmonella* in the

colon, can be made. This estimate should also be biologically plausible, therefore the likelihood of how this may contribute to the incubation period should be considered.

In order to effectively estimate the distribution of each parameter, the ideal level of information that should be extracted from each source includes a reliable empirical based measure of central tendency and of variation, such as the mean and standard deviation or the median and interquartile range. This level of information was mostly available in the human in-vivo studies and the in-vitro mouse experiment by Lowrie et al. Other studies provided a single estimate, with or without 95% confidence intervals. On occasions, the data provided will have to be fitted to models to derive predicted or expected data. This is most useful where reported data are sparse as in the experiments in steps 7 and 8 reporting the growth of bacteria in phagocyte and the death of phagocytes respectively.

Although parameter distributions are important, particularly in conducting sensitivity analysis of models where probability distributions are used to define the possible range of the value being analysed, sometimes these are unavailable as observed in this review. Where this is the case, an approach to sensitivity analysis could be identifying the lowest plausible value and the highest plausible value that would represent the range of the parameter value.

8.5 PUBLIC HEALTH AND RESEARCH IMPLICATIONS

This review has indicated that there are steps in the process of infection with substantial variability in their parameter values some of which are due to varied characteristics of study participants. The variation observed at each step in the infection process could result in an overall variation in the incubation period. In outbreak investigations or in planning public health response, individual variations and how this might influence the incubation period should be considered. Outbreak investigations should therefore be tailored to the characteristics of the population at risk as a standard approach may be ineffective considering individual variations.

The process of conducting this review identified gaps of evidence in the parameterisation of the process of infection. These evidence gaps can be explored using sensitivity analysis to observe how different parameter values will influence the incubation period. If the expected outcome is known, perhaps through a human experimental study, the parameter values can be altered to identify one that results in a similar outcome.

CHAPTER 9: WITHIN-HOST MATHEMATICAL MODELLING OF THE INCUBATION PERIOD OF *SALMONELLA* TYPHI

ABSTRACT

Mechanistic mathematical models are often employed to understand the dynamics of infectious diseases within a population or within a host. They provide estimates that may not otherwise be available. I have developed a within-host mathematical model in order to understand how the pathophysiology of *Salmonella* Typhi contributes to its incubation period.

The model describes the process of infection from ingestion to onset of clinical illness using a set of ordinary differential equations. The model was parameterised using estimated values from human and mouse experimental studies and the incubation period was estimated as 9.6 days. A sensitivity analysis was also conducted to identify the parameters that most affect the derived incubation period.

The migration of bacteria to the caecal lymph node was observed as a major bottle neck for infection. The sensitivity analysis indicated the growth rate of bacteria in late phase systemic infection and the net population of bacteria in the colon as parameters that most influence incubation period.

I have shown in this chapter how mathematical models can be used in understanding biological processes and estimating parameters of infectious diseases.

9.1 INTRODUCTION

Typhoid fever is a systemic infection caused by the bacteria *Salmonella Typhi*. It is endemic in developing countries; however, a sporadic disease in developed countries with infection occurring mainly in travellers returning from endemic regions (185). Community wide outbreaks associated with widely distributed products (203,249) or water supply (200,201,250) are commonly reported as well as point source outbreaks associated with dairy (189), red meat (208,251) and vegetables (197,252) have been reported.

Possible control measures that can be applied to typhoid fever depends on the population. In developing countries, this will involve provision of portable drinking water and hygienic food preparation (185). In developed countries, activities such as prompt investigation of outbreaks, identification of travel related cases and confirming cases as either sporadic or outbreak related, all of which require accurate knowledge of the incubation period, contribute to reducing the burden of disease. Furthermore, the efficacy of typhoid vaccine has been tested in experimental and field studies.

The incubation period of typhoid fever is defined as the time between exposure to the pathogen and the onset of clinical illness. Individual variation in the incubation period due to host and pathogen characteristics mean that this parameter is reported as a distribution rather than a single estimate. In addition to understanding the pathophysiology of the disease, accurate knowledge of the distribution of incubation period is important for surveillance, outbreak investigations and conducting epidemiological and ecological studies (48). Available reports on the incubation period of *S. Typhi*, vary extensively. Some community wide outbreaks have reported incubation periods with a mean of 19 days (204) or a range of two to eleven days (253) and some point source outbreaks have reported ranges of five to twelve days (189) and one to twenty-eight days (210).

International health organisations such as WHO and CDC (63,188) report the incubation period as ranging from three to sixty days and three to thirty days respectively, possibly in an

attempt to include the possible extremes of the distribution. These estimates are probably based on observations from a limited number of studies and factors influencing the distribution are unknown. Understanding the distribution of incubation period and the influencing factors may be possible by undertaking a systematic review of observational or experimental studies, however mathematical models have the potential to provide a more complete description and identify influencing factors not detected using other research methods.

Mathematical models have been employed in understanding areas of infectious diseases such as the spread of a pathogen within a host (77) and the immune process of the human body (78). They can also be used to estimate parameters like incubation period and dose response or predict outcomes by fitting data to the developed models. The process of model formulation helps to increase understanding of the pathophysiology, clarify and document assumptions and data gaps. It is also useful in assessing the sensitivity of the model output to changes in parameter values in an attempt to identify influencing factors.

Based on the literature review conducted in chapter 8, which summarised the pathophysiology of *S. Typhi*, I developed a compartmental model that described the process of infection from ingestion of a contaminated liquid meal to onset of clinical illness using a set of ordinary differential equations. Each equation represented the transition from one compartment to the next and estimates the duration of each step. The model was parameterised using estimated values identified in chapter 8. Observed values from the model output was compared to expected values identified from data of a human experimental study. I also conducted a sensitivity analysis to identify parameters or variables that influence the distribution of incubation period.

9.2 MODEL FORMULATION

Based on the findings from the literature review (Chapter 8), the summarised pathophysiology was translated into mathematical equations and a compartmental model was developed. This represented the transfer of bacteria per unit time from initial ingestion to onset of clinical symptoms (mouth to secondary bacteraemia) (Figure 30). Ordinary differential equations were used to describe the rate of change of bacteria in each compartment which included a combination of the processes: migration, replication or clearance. The model was solved using the R statistical package (127) .

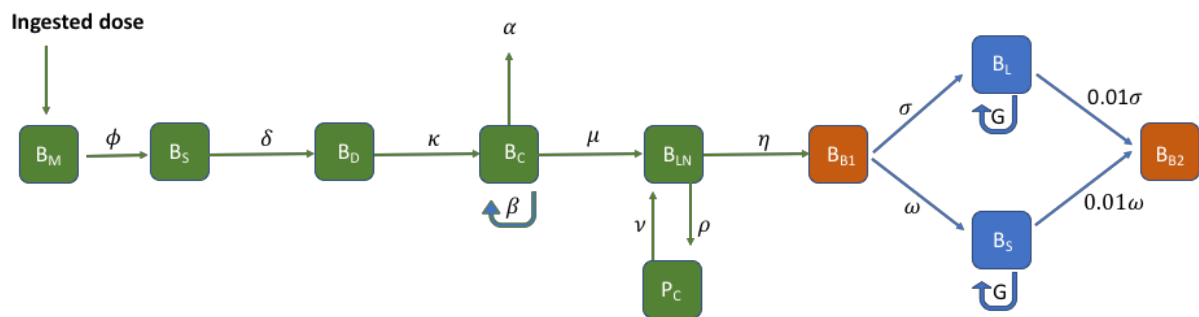


Figure 30 Flowchart of model showing the mathematical representation of the infection process

9.2.1 Invasion of the GI organs

With the understanding of the pathogenesis of *S. typhi* in the GI tract and the different steps involved in this phase of GI invasion, a set of ordinary differential equations that describe the process and estimate the duration of each stage of infection as the bacteria leaves one compartment to the other was defined. The bacteria enter into the stomach within seconds of ingestion and I assumed that about 95% of the bacterial population in the stomach enter into the small intestine. I also assumed that in the small intestine, there is no lag time before bacteria begins to transit, and in the colon, bacteria sheds at one and a half times the replication rate.

The rate of migration from the mouth to the colon is defined as:

$$\dot{B}_M = -\phi B_M \quad (1)$$

$$\dot{B}_S = \phi B_M - \delta B_S \quad (2)$$

$$\dot{B}_{D,1} = \gamma \delta B_S - \kappa_1 B_{D,1} \quad (3)$$

$$\dot{B}_{D,i+1} = \kappa_i B_{D,i} - \kappa_{i+1} B_{D,i+1} \quad (4)$$

$$\dot{B}_C = \kappa_7 B_{D,7} - (\mu + \alpha - \beta) B_C \quad (5)$$

where:

ϕ is the transit time from mouth to stomach;

δ is the gastric emptying rate;

γ denotes the fraction of bacteria that successfully migrate to the duodenum;

$\kappa_i = \kappa/7$ where κ is the mean transit time through the small intestine and $i = 1, \dots, 7$ are the 7 compartments in the small intestine reported by Yu et al (254);

β represents the rate of replication in the colon;

α is the rate of bacteria that is shed from the colon through faeces;

μ is rate of bacterial migration;

B_M is the bacterial population in the mouth;

B_S is the bacterial population in the stomach;

B_D is the bacterial population in the small intestine;

B_C bacterial population size in the colon.

9.2.2 Invasion of caecal lymph nodes and phagocytes

$$\dot{B}_{LN} = \mu B_C + \nu C P_C - (\rho + \eta) B_{LN} \quad (6)$$

$$\dot{P}_C = \theta \rho B_{LN} - \nu P_C \quad (7)$$

Therefore, the approximate population of bacteria available in cecal lymph node and phagocytes is:

$$T = B_{LN} + C P_C \quad (8)$$

where:

C denotes the number of bacteria in the phagocyte at time of rupture

P_C represents the number of invaded phagocytes

ν represents the rate of phagocyte rupture

ρ denotes the rate of phagocyte invasion

θ denotes the fraction of bacteria infecting phagocytes

η denotes the flow rate of the lymph to blood

T represents total population of bacteria in both cecal lymph nodes and phagocytes

9.2.3 Onset of primary bacteraemia

Using the evidence available from the mice model and the description of the systemic phase of *S. Typhi* infection, the onset of primary bacteraemia and the transfer of bacteria into the organs was described as:

$$\dot{B}_{B1} = \eta B_{LN} - (\omega + \sigma) B_{B1} \quad (9)$$

where: ω and σ are transfer rates of bacteria from blood to the spleen and liver

9.2.4 Invasion and replication of bacteria in systemic organs

Net bacterial growth in systemic infection is defined as:

$$\dot{B}_S = \omega B_{B1} + \tau B_S - (\Upsilon\omega) B_S \quad (10a)$$

$$\dot{B}_L = \sigma B_{B1} + \tau B_L - (\Upsilon\sigma) B_L \quad (10b)$$

where:

Υ represents the reduction in flow back to the blood

τ represents the net growth of bacteria in organs during the invasion and replication phases

and so $\tau = \tau_1$ for early times ($t < t_1$) and $\tau = \tau_2$ otherwise

B_S and B_L denotes the bacterial population in the spleen and liver

9.2.5 Onset of secondary bacteraemia and clinical symptoms

$$\dot{B}_{B2} = (\Upsilon\omega) B_S + (\Upsilon\sigma) B_L \quad (11)$$

9.3 MODEL PARAMETERISATION

The model has 17 parameters (the initial number of bacteria ingested is not included as a parameter) and parameter values were derived from the literature review in chapter 8.

Reviewing the available evidence for the various transit times and rates, I identified, calculated and estimated values for the model parameters. Where possible, values indicating duration or proportion were converted to rates. Bacteria transit times in the gastrointestinal tract and flow rates in the lymph and blood were estimated using the identified human experimental studies. At cellular level, I assumed that the replication and death of bacteria and their interactions with macrophages are similar in mice and humans, hence, several mouse models of *Salmonella typhimurium* (*S. typhimurium*) infection was used to describe cellular infection. A list of the estimated parameter values to be included in the model is available in Table 24.

9.3.1 Gastric emptying rate (δ)

Following ingestion, the bacteria transits through the mouth and arrives in the stomach within a few seconds (ϕ) (this parameter is not tested for sensitivity as it is so short a duration).

According to Bennink et al, the liquid emptying rates for both males and females are similar (2.34 %/min and 2.33 %/min) (220), hence, I decided on a single emptying rate for both genders of 2.3 %/min. I calculated the average emptying time using the equation:

$$p(t) = 1 - e^{-t/\delta} \quad (12a)$$

where: $p(t)$ is the proportion emptied per minute (0.023) and t is time (1 minute), therefore;

$$\delta = \frac{-t}{\ln(1-p(t))} \quad (12b)$$

$$\delta = \frac{-1}{\ln(1-0.023)} = 42.9$$

It will therefore take an average of 43 minutes for gastric content to be emptied at a rate of 1.40 per hour.

Table 24 Values, range and distribution of parameter values used in the sensitivity analysis

Parameter	Description	Parameter value	Unit	Range	Reference	Distribution	Rationale for values, range and distribution
B_0	Ingested dose		Counts				
ϕ	Transit rate through the mouth		Rate/hour				
δ	Gastric emptying rate	1.4	Rate/hour	SD $\pm$ 0.7	Bennink et al (220)	Normal	Multiple published evidence available. Data chosen as authors reported overall mean ($\pm$ SD) emptying rate for liquids.
γ	Fraction of bacteria entering the small intestine	0.95	Proportion	Range: 0.8 - 1		Uniform	Value is assumed to be 95% as literature indicates almost all bacteria enter the SI. The values of the range were chosen to include an upper value that allows for 100% of bacteria to go through considering the food vehicle of the experimental study.
κ	Small intestine emptying rate	0.3	Rate/hour	SD $\pm$ 0.1	Yu et al (254)	Normal	Multiple published evidence available. Value chosen as data was estimated from over 400 human transit data. Evidence suggests that transit time distribution is not log-normal.
α	Rate of bacteria shedding from the colon	0.36	Rate/hour	Range: 0.1 – 1.9		Uniform	Value was selected based on best guess estimate. Expecting that some proportion of bacteria must be shed from the colon, the lower limit of the range

							was set at 0.1. The upper limit was set to be below the replication rate, because if shedding occurs faster than replication, infection will be cleared.
β	Replication rate in the colon	0.43	Rate/hour	Range: 0.1 - 2.1	Knodler et al (229)	Triangular	Value based on published evidence. Parameter value is the average doubling time. Upper limit represents the replicating rate of the hyper-replicating cells reported by Knodler et al. Lower limit is a best guess estimate of the slowest possible replication rate. As no distribution is available in the literature, a triangular distribution is selected to include the min and max values.
μ	Migration rate to the caecal lymph node	1.25×10^{-8}	Rate/hour	Range: 0 to 5.5×10^{-8}	Kaiser et al (227)	Triangular	Published evidence reports mean value and 95% CI. The corresponding SEM from the 95% CI was used to derive the SD in order to extend the range around the mean. A triangular distribution was selected to include the mean and $+1.96 \times \text{SD}$ as the upper limit of the range and lower limit was censored at 0 as a negative migration rate would result in negative model outputs.

θ	Fraction of bacteria replicating in a phagocyte	0.3	Proportion	Range: 0.1 to 0.6		Triangular	Parameter value is the average number reported across three experiments. From the range of values reported in the experiments, the lower and upper limit of the range were selected to be minimum and maximum value reported in the experiments, expecting that the fraction of bacteria available to replicate cannot be less than 10% or more than 60%.
ρ	Rate of phagocyte invasion	0.09	Rate/second	Range: 0.018 to 0.162	Gog et al (232)	Triangular	Two values were reported in the literature from a physical model and a mathematical model. The value from the physical model was selected as the more biologically plausible of the two as it did not depend on the multiplicity of infection (MOI), which is difficult to ascertain during infection. The upper limit of the range was selected to be similar to the value reported in the mathematical model as the maximum possible rate.
ν	Rate of phagocyte rupture	0.41	Rate/hour	Range: 0.20 - 0.60	Monack et al (238)	Triangular	The values of the range were selected based on best guess estimates. If the rupture rate is below 0.1 or close to 1, infection might not occur. If the phagocytes rupture too slowly, there will

							be more bacterial death occurring; and if the phagocyte rupture to quickly, there will be insufficient replication of bacteria within the phagocyte.
C	Number of bacteria in phagocyte at rupture	4.1	Counts	Range: 1 - 10	Calculated	Triangular	The range of values selected would allow for a lower but longer tail, while still centred around the peak value.
η	Flow rate from lymph to blood	0.025	Rate/hour	Range: 0.005 to 0.315	Frietas et al (242)	Triangular	Parameter estimate is selected from the report of Frietas et al. The upper limit of the range is the value reported by Alexander et al which would represent the fastest flow rate.
σ	Mean blood flow rate in portal vein	7.9	Rate/hour	SD $\pm$ 2	Brown et al (245)	Normal	Values based on published evidence. It is assumed that the flow rate follows a normal distribution with the reported standard deviation around the mean.
ω	Mean blood flow rate in splenic artery	2.4	Rate/hour	SD $\pm$ 0.4	Sato et al (247)	Normal	Values based on published evidence. It is assumed that the flow rate follows a normal distribution with the reported standard deviation around the mean.
τ (represented by τ_1 and τ_2)	Net growth rate in early and late stages of systemic colonization	-1.04 in early phase 0.09 in late phase	Rate/hour	Range in early phase (τ_1): -2 to 0 Range in late phase (τ_2): 0 to 1	Grant et al (77)	Triangular	In early phase, bacterial death is higher than replication. The range is set so that at the upper limit, death rate is equal to replication rate, and at the lower limit, death rate is twice the replication rate.

							In late phase, bacterial death is negligible, thus, the range is set such that at the lower limit, replication rate is equally negligible and at the upper limit, the replication rate is one magnitude higher.
γ	Magnitude of reduction in flow rate from systemic organs back to blood	0.01	Proportion	0.001 to 0.1	Grant et al (77)	Triangular	Best guess estimate that proportional change in flow rate can be as low as 0.1% and no higher than 10%.

9.3.2 Fraction of bacteria migrated to duodenum (γ)

For the purpose of this model, I assumed that gastric acid was neutralised and the altered gastric pH increases the likelihood of the ingested bacteria to pass through the stomach unrestricted. Thus, I estimated that about 95% (0.95) of the ingested bacteria migrated to the duodenum.

9.3.3 Small bowel transit time (κ)

The compartmental transit model developed by Yu et al can be used to characterise the variation in the population (254). The model assumes that the gastric content flow through the small intestine by passing through a series of segments/ compartments, where each compartment has different food volumes and flow rate but equal transit times. The mean transit time can be derived, and the ideal number of compartment (N) can be determined from the compartmental analysis by adding and subtracting compartments until the residual sum of squares (SSE) becomes small. From their analysis, Yu et al identified a seven-compartment transit model as the best model to describe the flow process (226).

Reproducing the analysis from Yu et al (226), the values of N with the smallest SSE were 6.9 and 7.0 and the corresponding SSE were 10.8 and 11.1 respectively. Although the SSE for 6.9 was the smallest, inputting non-integers into the model could not be justified as it is not possible to have a fraction of a model compartment. Hence, I opted for seven compartments and a mean transit time of 199 min (κ), which was the same value reported by Yu et al. This resulted in an emptying rate of 0.30 per hour.

9.3.4 Bacteria in the colon (β , μ and α)

The average doubling time of *Salmonella* in the colon, according to Knodler et al (229) was 95 min. From the doubling time formula

$$\log 2 = rt \quad (13)$$

where r and t are the replication rate and doubling time (in hours) respectively; I solved for r as $r = \frac{\log 2}{t}$. The resulting replication rate was $\beta = 0.43$.

According to the combined experimental study and mathematical models undertaken by Kaiser et al (227), the resulting probability of migration in the first 24 hours was 3×10^{-7} . Therefore, the probability of a bacterium migrating in one hour is $\mu = 1.25 \times 10^{-8}$.

Considering the high replication rate of bacteria in the colon and the low probability of migrating to the caecal lymph node, more bacteria will be shed through faeces compared to the quantity that continue towards systemic infection.

However, for gut invasion to be sustained and for infection to progress, the replication rate of bacteria in the colon should be higher than the rate it sheds through faeces. In order to avoid total bacterial clearance, the net bacterial population should to be greater than 1.

Considering that the average doubling time of bacteria is approximately 95 min, it would be expected that for gut invasion to be sustained, the clearance of bacteria would occur at a slower rate and the time it takes bacteria to shed from the colon would be nearly twice as long as the doubling time. In the absence of evidence in the literature, a shedding time that is 1.8 times the doubling time is the best guess estimate, allowing the net population of bacteria in the colon to be greater than 1, yet not resulting in a very slow rate that may unreasonably shorten the incubation period. This results in a shedding time of 171 min and a corresponding shedding rate per unit time of 0.35 per hour.

9.3.5 Rate of phagocyte invasion/engulfment (ρ)

Based on the physical model conducted by Gog et al (232), the rate of phagocyte invasion was 2.5×10^{-5} per second per bacterium or 0.09 per hour.

9.3.6 Proportion of bacteria replicating in a phagocyte (θ)

Dissel et al (255) carried-out experiments and developed a model to estimate the intracellular killing of bacteria using phagocytes from three types of mice. In both the

experiments and the model, the number of viable bacteria in the phagocyte decreased after 90 mins in all three types of macrophages. In the experiment, the percentage of viable bacteria decreased to 69%, 25% and 13% while from the developed model, the number of bacteria decreased to about 20% in two types of macrophages and to 50% in the third (255). Based on the experiment and model developed by Dissel et al (255), the average proportion of viable bacteria available in the cell after phagocytosis in the three types of mice was approximately 30%.

9.3.7 Growth rate of bacteria in phagocyte

Extracting the experimental data from Forest et al (236) for the increase in intracellular bacterial population, I derived the replication rate of intracellular bacteria. I fitted the extracted data to a logistic replication model (Figure 31):

$$P = \frac{CB_0}{B_0 + (C - B_0)e^{-rt}} \quad (14)$$

where P is the number of bacteria in the phagocyte after replication, B_0 is the initial number of bacteria in the phagocyte, and C , r , and t represent the maximum value (the “carrying capacity”), exponential growth and time respectively. The predicted intracellular growth rate was 0.3 per hour in the first six hours after infection which results in a doubling time of 2.3 hours.

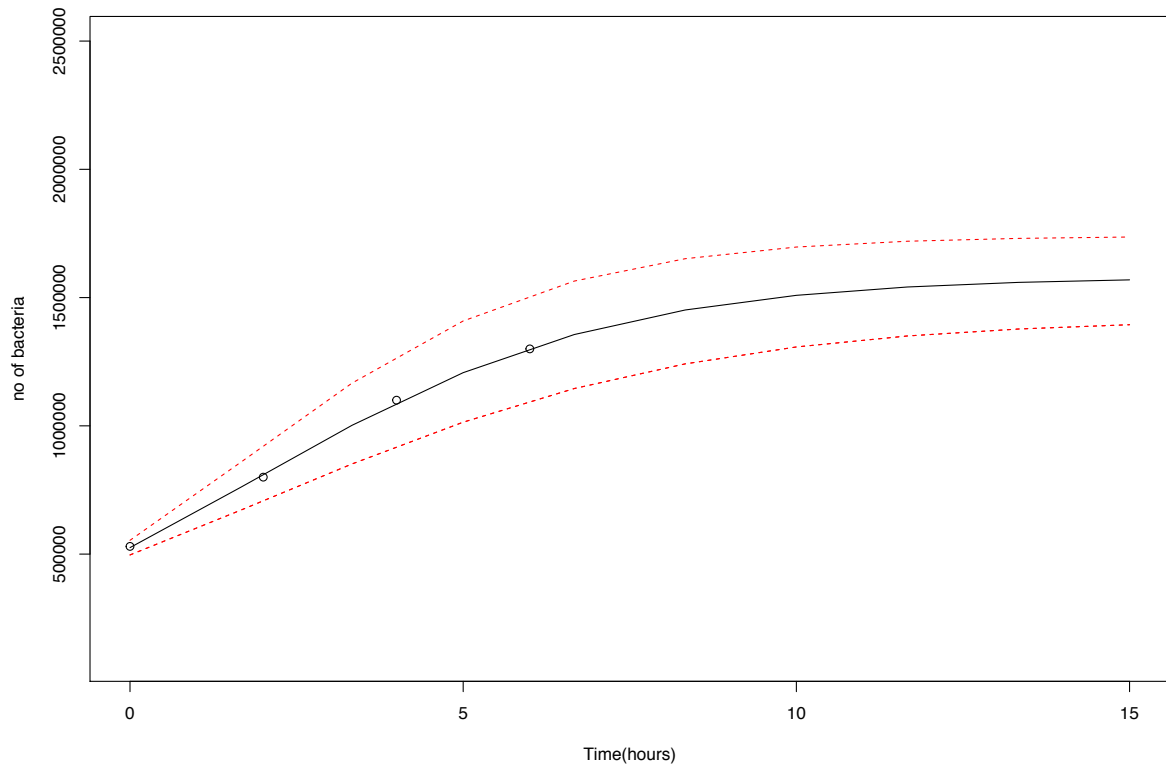


Figure 31 Data from Forest et al fitted to replication model in estimating the growth rate of bacteria in phagocytes with 95% confidence intervals

9.3.8 Rate of phagocyte rupture (ν)

Data from Monack et al (238), on the proportion of dead phagocytes at different time intervals after infection, were extracted. The median rupture time, which is the time it will take for 50% of the phagocytes to rupture, is chosen to be representative of the average rupture time. I fitted the data to a logistic model (after checking the fit against one using a lognormal distribution):

$$f(x) = \frac{1}{1 + e^{-k(x-x_0)}} \quad (15)$$

The resulting intercept and gradient were 4.59 and -0.03179 respectively. Inputting these parameters into equation 15 and solving for x resulted in a value of 144.38 (4.59/-(-0.03179)), which represented the 'median rupture time'. Hence, the time it took for 50% of phagocytes to rupture was 144.38 min or 2.41 hours. This corresponds to a rupture rate of 0.41 per hour.

9.3.9 Population of bacteria in phagocyte at rupture (C)

The number of bacteria in the phagocyte at time of rupture depends on the rate of intracellular bacteria replication and rate of phagocyte rupture. Using the replication rate predicted from the model in equation 14, and the median rupture time from the model in equation 15, I derived the number of bacteria in the phagocyte at the time of rupture. According to an experiment by Mitsuhashi et al (256), when phagocyte cells and bacteria were incubated at a ratio of 1 to 10, the proportion of infected phagocyte cells was 30% of which 80% ingested one to two bacteria. I therefore assumed that at the time of phagocytosis, each cell would contain a maximum of two bacteria. Inputting the derived doubling time of 2.3 hours and the required duration of 2.4 hours, which is the median rupture time, into the formula $B(2)^{t/d}$ where B is the initial number of bacteria available, t is the period of interest and d is the doubling time, it can be estimated that the number of bacteria in the phagocyte at rupture would be 4.1.

9.3.10 Draining rate of lymph from the caecal lymph node (η)

In order to estimate the flow rate out of the caecal lymph node, I assumed that bacteria will drain from the caecal lymph node at the same rate that lymph flows through the intestinal lymphatic vessel. The vessel has a flow velocity of 3400 μ m/sec and a corresponding flow rate of 14mm³/sec (242), equivalent to 0.014ml/sec. The lymph vessels in the human body have a total capacity of approximately 2L (242), therefore the intestinal lymphatic vessel rate is 0.025/hr.

9.3.11 Transfer rate into systemic organs (σ and ω)

In order to estimate the transfer rate of bacteria into the organs, I assumed that bacteria will travel into organs at the same rate as blood flow in the portal vein and splenic arteries and that equal proportion of bacteria enter into both organs. The mean blood velocity in the

portal vein was reported to be 13.9 ± 4.49 cm/s and the resulting mean volume flow was 662 ± 169 mL/min (245). Considering that the total blood volume of a typical adult is 5 litres (5,000mL), the transfer rate into the portal vein is therefore 0.13/min.

According to Sato et al (247), the mean blood velocity in the splenic artery was 18.7 ± 4.2 cm/s and the resulting mean blood flow rate was 179 ± 37 mL/min. The transfer rate into the spleen was calculated as 0.04/min.

9.3.12 Net growth of bacteria in systemic organs (τ)

According to the mouse model and accompanying mathematical model developed by Grant et al (77), the bacterial population in the organs slowly decreases in the first six hours with a doubling time of 1.7 hours and a half-life of 1.1 hours. The replication/birth rate was calculated using the formula in equation (13), and the death rate was calculated using the half-life formula:

$$\log 0.5 = rt \quad (16)$$

The resulting birth rate and death rate were 0.41 and -0.63 respectively. Thus, the net growth rate for both spleen and liver was -1.04 in the first seven hours post primary bacteraemia during the invasion phase.

In the replication phase of exponential bacterial growth, the doubling time decreases to eight hours (77), and with negligible death rate, results in a net growth rate of 0.09.

9.3.13 Transfer rate of bacteria from systemic organs into the blood and onset of secondary bacteraemia

Based on the mouse model by Grant et al, the relative rate of bacterial transfer out of the systemic organs was about one per cent (1%) of the bacterial transfer rate from blood to the organs (σ and ω). Therefore, the transfer rates from the liver and spleen are 0.0013/min and 0.0004/min respectively.

9.4 MODEL ANALYSIS

From the model, I observed three critical stages for infection mathematically, and perhaps biologically, including: the net population of bacteria in the colon (determined by $\beta > \alpha$), the bacterial population during phagocytosis (determined by $\theta C > 1$) and the growth rate in late systemic infection (determined by $\tau_2 > 0$). Most compartments in the model act to transit the bacterial load to other locations in the body or remove bacteria. If these 3 conditions on parameters are not met then bacterial population will reduce over time and result in no infection.

9.5 SOLVING THE MODEL AND INTERPRETATION

The initial state of each compartment was defined. B_0 was defined as the initially ingested dose of bacteria with a concentration of 10^4 . Therefore, the initial conditions were set such that all states are zero except $B_M(0) = B_0$.

The transfer of bacteria through the stomach, small intestine and colon is represented in Figure 32. Gastric contents are completely emptied within five hours of ingestion. Transit through the small intestine follow a gamma distribution and using the seven sub-compartments of the small intestine (226) in the model resulted in a slight delay before the concentration of bacteria in the colon begins to increase (Figure 32). The migration of bacteria to the CLn and the replication of bacteria in the phagocyte is represented in Figure 33.

The invasion of the caecal lymph node to the onset of secondary bacteraemia is represented in Figure 34. At the early stages of systemic infection, the concentration of bacteria in the liver was higher than that in the spleen and blood, but as the infection progressed, bacterial concentration in the spleen and blood increased.

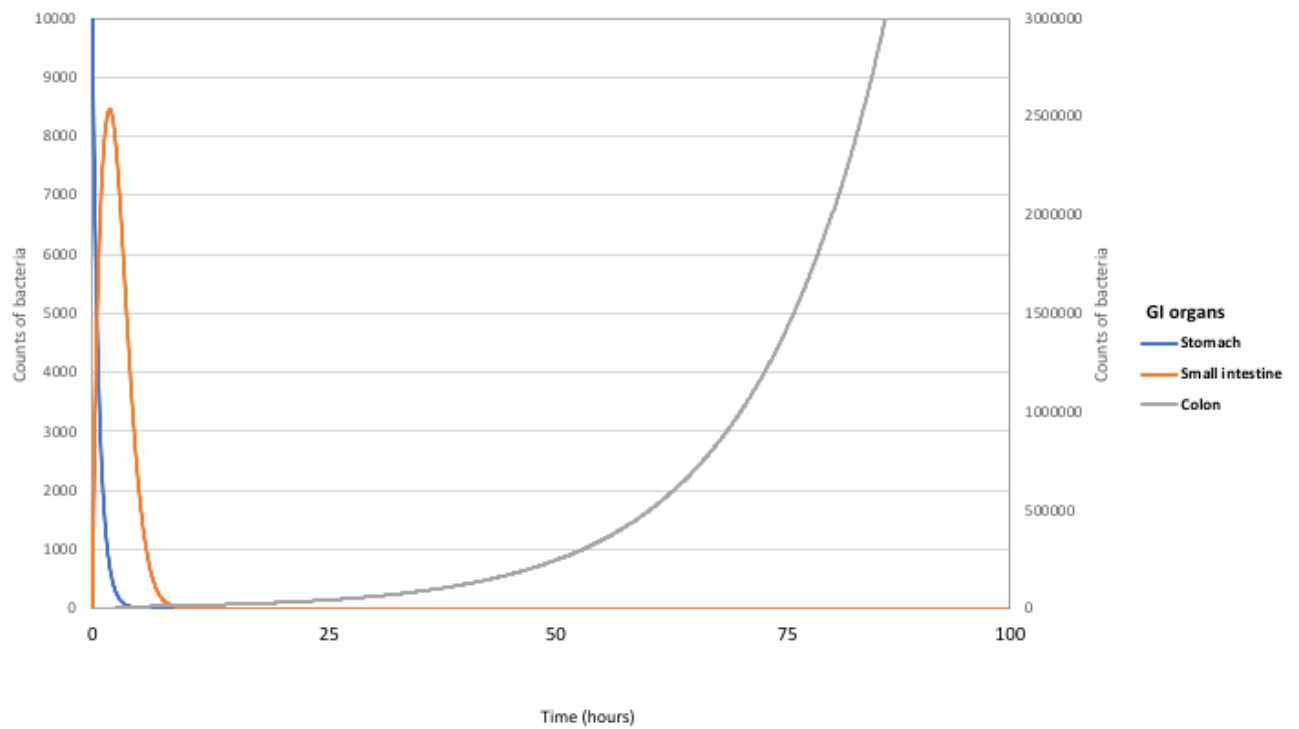


Figure 32 Model output showing the initial GI phase of the infection process in the within-host mathematical model

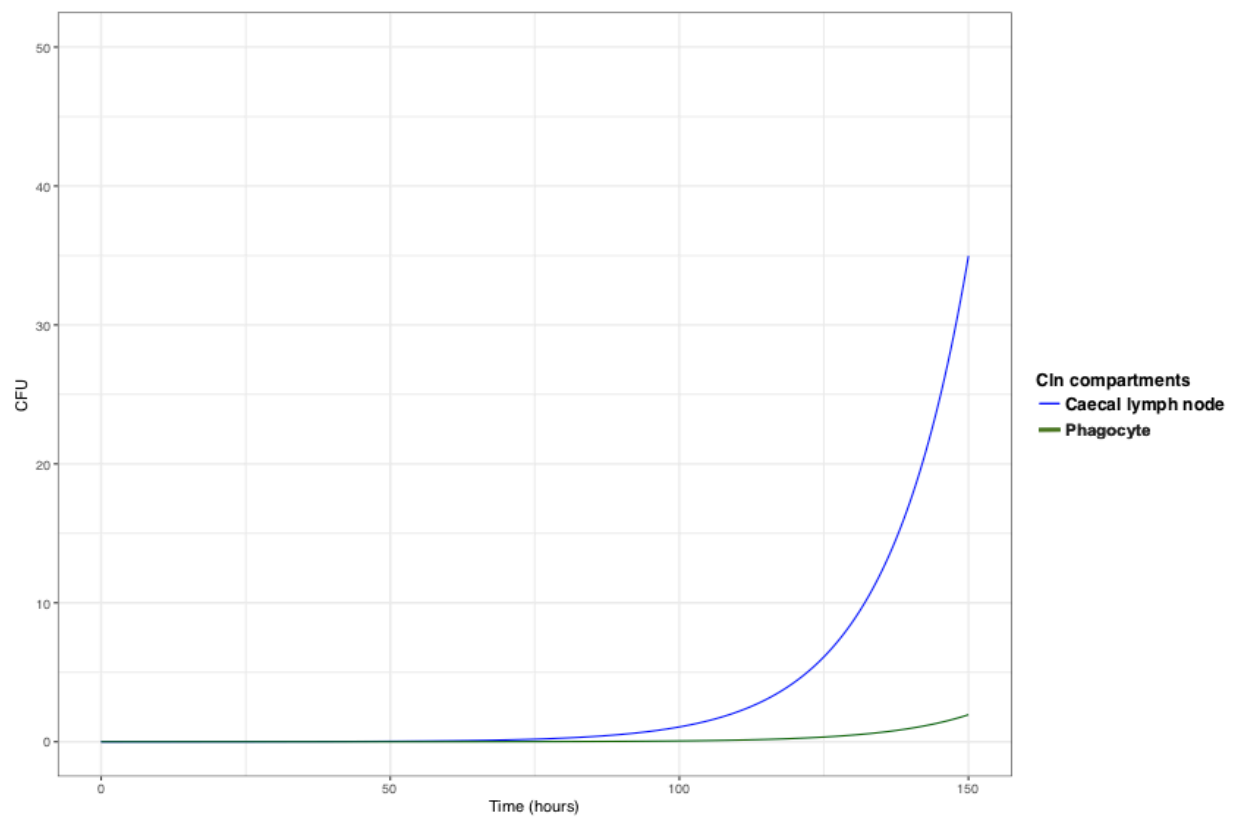


Figure 33 Model output showing the involvement of the CLn and phagocyte of the infection process in the within-host mathematical model

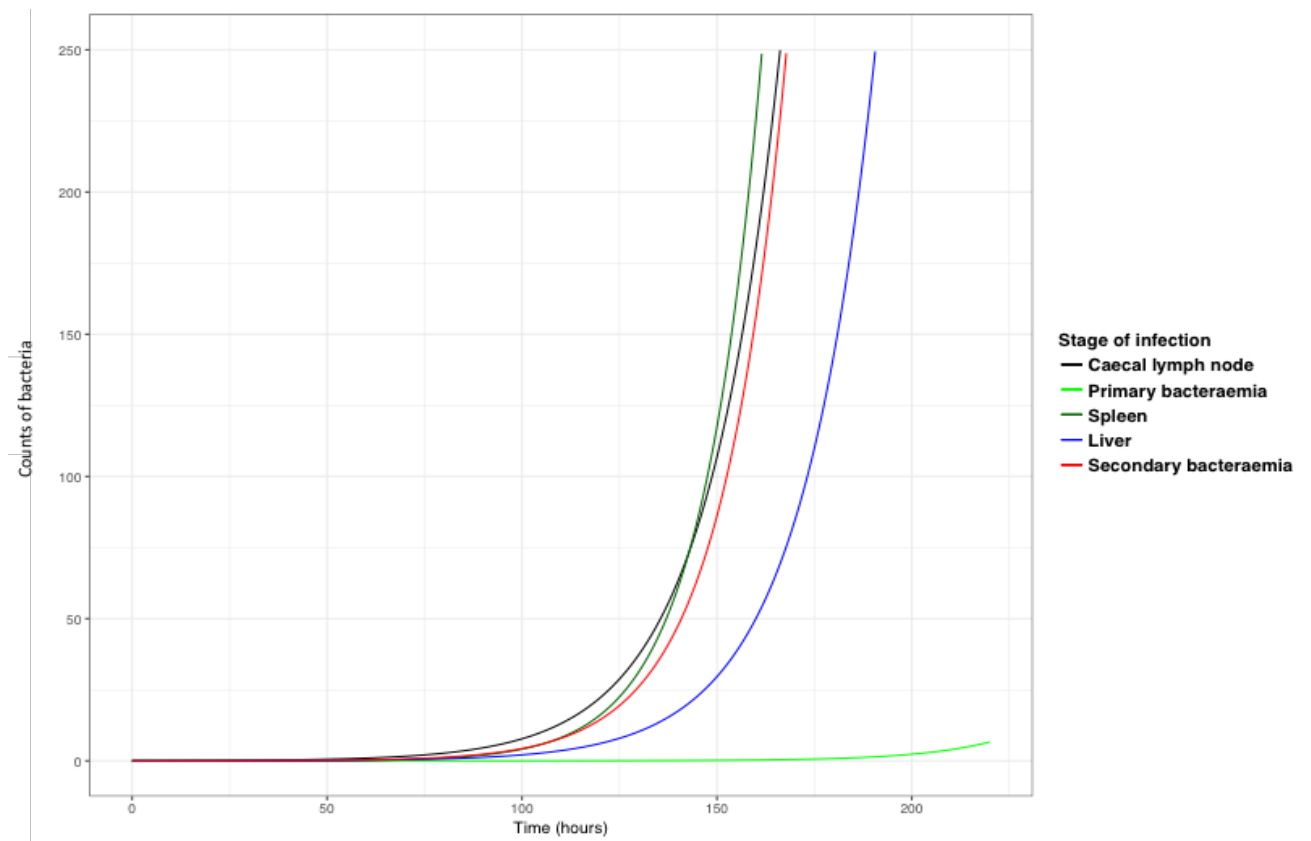


Figure 34 Model output showing bacteria in the CLn to secondary bacteraemia of the infection process in the within-host mathematical model

9.6 ESTIMATING INCUBATION PERIOD

Using the experimental study conducted by Waddington et al (65), the incubation period from this model output was estimated. In the experimental study, two sets of participants were challenged with 10^3 and 10^4 and monitored until typhoid diagnosis (TD). For cases with a dose of 10^4 , the median incubation time to TD was eight days (IQR 6 – 9 days).

Quantitative blood culture performed at time of TD showed median bacterial loads of 1.1 colony forming units (CFU)/ml (IQR 0.4 – 2.1).

From the mathematical model, the bacterial concentration observed at eight days is 345 CFU. As the total volume of blood is 5,000ml, the resulting bacterial load from the model at eight days was 0.07 CFU/ml. The bacteria load in the model during secondary bacteraemia reached the level reported from the experimental study at 9.6 days. The bacteria concentration in the blood at this time was 5513 CFU, which is equivalent to 1.1 CFU/ml.

9.7 EFFECT OF DOSE ON INCUBATION PERIOD

In order to examine the effect of dose on the model output, I altered the value of the initially ingested dose of bacteria such that B_0 was defined as a concentration of 10^3 . From the experimental study, cases with a dose of 10^3 reported a median incubation time of nine days to TD (IQR 6.5 – 13 days); and the quantitative blood culture performed at typhoid diagnosis showed median bacterial loads of 0.5 CFU/ml (IQR 0 – 1.2).

The output of the mathematical model showed the bacterial concentration at nine days to be 0.03 CFU/ml (189 total CFU count); and the bacterial load reached the level observed in the experimental study at 10.5 days, one day longer than the higher dose of 10^4 (Table 25). In both the experimental study and the mathematical model, an increase in ingested dose resulted in an increase in bacterial concentration in the blood by a ratio of about 2.1, leading to a shorter incubation period (Table 25).

Table 25 Values from mathematical model and experimental study showing effect of dose on incubation period

Challenge dose & mathematical model initial state	Experimental study (Waddington et al)		Mathematical model	
	Time to typhoid diagnosis (median (IQR))	Bacterial count at typhoid diagnosis (median (IQR))	Time to reach typhoid diagnosis bacterial levels (0.5 & 1.1 CFU/ml)	Bacterial count at time of typhoid diagnosis (9 & 8 days)
10^3	9 days (6.5 – 13)	0.5 CFU/ml (IQR 0 – 1.2)	10.5 days	0.03 CFU/ml (189 total CFU count)
10^4	8 days (6 – 9)	1.1 CFU/ml (IQR 0.4 – 2.1)	9.6 days	0.07 CFU/ml (345 total CFU count)

9.8 UNCERTAINTY SAMPLING AND SENSITIVITY ANALYSES

The parameter values inputted in the model have some degree of uncertainty due to factors such as natural variation between individuals or measurement error in experiments.

Therefore, an uncertainty analysis to quantify the degree of uncertainty in the parameter estimates and then a sensitivity analysis to explore the effect of changes in the model parameters on the incubation period were conducted. The sensitivity analysis was conducted in a two-step process: first determining the effect of each parameter individually on the incubation period and identifying the most sensitive parameters; then determining the combined effect of all identified sensitive parameters on the incubation period. I used a computational modelling software called Sampling and Sensitivity Analysis Tool (SaSAT) developed by Hoare et al (257) for both the uncertainty sampling and the sensitivity analysis.

9.8.1 Uncertainty sampling

Latin Hypercube Sampling (LHS) method, a form of Monte Carlo simulation, was used in conducting the uncertainty analysis. It uses stratified sampling (258) to randomly select parameter values within the given parameter distributions that result in an unbiased estimate of the average model output (259). For each parameter value, I specified a range from which the random values were selected (Table 24). Some of the parameter values had corresponding measures of variation such as standard deviation or 95% confidence interval, hence the range specified was based on empirical evidence. For other parameter values, the range was selected based on best available evidence from the literature or best guess estimates based biological expectations. I also defined the underlying distribution of each parameter value from which the random samples were selected. A triangular distribution was mostly chosen in the absence of evidence for other type of distribution, and this included a peak value and plausible minimum and maximum values. Due to the triangular shape of the distribution, more random samples were selected around the peak value. A uniform distribution was selected when there was neither evidence of other type of distribution, nor evidence supporting the parameter value. In this case, minimum and maximum values are selected to give a range which could include the actual parameter value. The uniform sampling from this range, due to the uncertainty of the parameter value, produces a uniform distribution. The parameter estimates, range and underlying distribution used to derive the random parameter values are summarised in Table 24.

In the first step of the uncertainty sampling, a hundred random samples were generated for each parameter, and in the second step, a thousand random samples were generated. Figure 35 shows the probability distribution functions of 1,000 randomly generated samples for each parameter displaying the shape of the distribution and the range of the values selected.

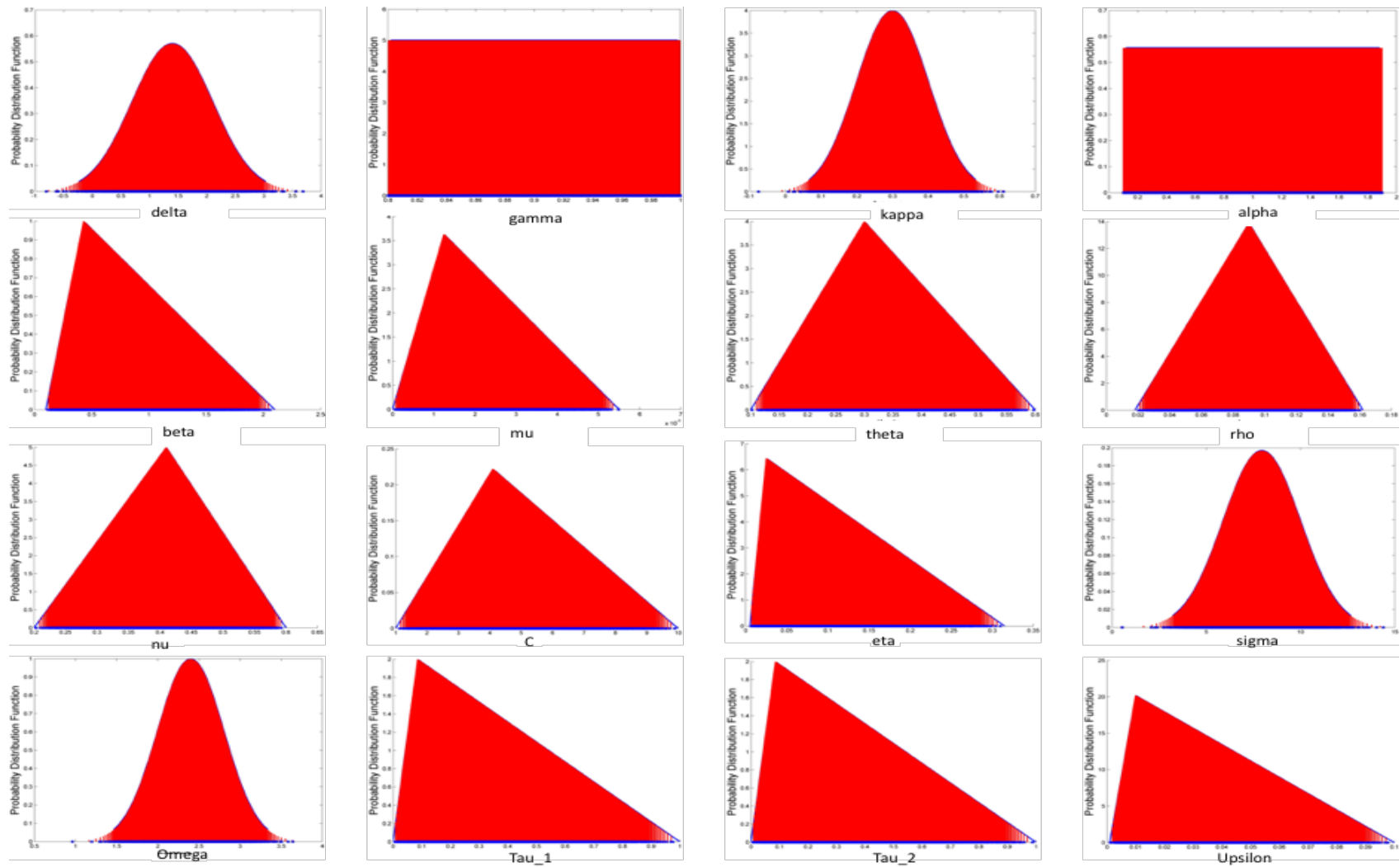


Figure 35 Probability distribution plots showing range and shape of parameter values used in the sensitivity analysis of the within-host mathematical model

9.8.2 Sensitivity analysis

In order to examine the sensitivity of each parameter independently, the single parameter models were run with the values of the parameter of interest varying according to the samples generated by the LHS, while the values of the other parameters were fixed. For example, when examining the sensitivity of δ , the model was run with varying values of δ and fixed values for the other parameters. Using the 100 random samples generated by the LHS, 100 reiterations were run for each parameter considered in the sensitivity, resulting in sixteen sets of 100 model outputs (the transit time for the mouth to stomach was not included in the sensitivity analysis as this was a short duration). Scatter plots for each parameter, plotting the parameter value against the resulting incubation period, were generated (Figures 36 & 37). Comparing the output of all simulated models, six parameters were identified as the most sensitive including: β , α , μ , C , θ , τ_2 . This may be expected from consideration of the equations themselves as discussed in the model analysis section above as these contribute to the bacterial growth rather than simply tracking transit of bacteria around body. Of these parameters is the most uncertain as it is the only parameter without supporting evidence from the literature.

Two simulated values of delta were negative, but these are not biologically feasible and so can be rejected. The exponential phase on the 'alpha' curve indicates the period where $\alpha < \beta$, thus, an exponential growth in the bacterial population will be observed resulting in the steep rise in the incubation period. Likewise, the exponential decline on the 'beta' curve indicates that $\beta > \alpha$. At the tail end of the curve, we observe the point of diminished return, whereby no matter how much greater β is, the increasing bacteria in the colon cannot shorten the incubation period any further (Figure 36). Values of $\beta < \alpha$ will cause net reduction so fractional bacteria would enter the CLn, hence, these extreme values are perhaps less biologically feasible.

The correlation of the bacterial growth in late phase (τ_2) is independent of the decay in the early phase (τ_1), such that regardless of the level of decay experienced in the latter, infection is most likely to occur if bacterial growth occurs in the late phase (Figure 37). Hence, τ_2 is one of the determinants of infection.

In the second step of the sensitivity analysis, 1,000 reiterations of the model were ran using 1,000 randomly generated samples of the six sensitive parameters. Using the output from the simulated models, I conducted a sensitivity analysis by deriving the partial rank correlation coefficients (PRCC) and the corresponding p-values (Table 26). This provides a measure of the strength of the association between the six sensitive parameters and the incubation period. The parameters with the strongest correlation with incubation period included τ_2 , α and β with correlation coefficients of -0.7, 0.6 and -0.6 respectively (Figure 38). Growth rate in late phase systemic infection (τ_2) was most strongly correlated with the incubation period, where an increase in the growth rate significantly reduced the incubation period (Figure 39). With the exception of C (number of bacteria in phagocyte at rupture), all parameters were significantly associated with the incubation period (Table 26).

Table 26 Partial rank correlation coefficients of model parameters

Parameter	Description	Correlation coefficient	p-values
τ_2	Growth rate in late stage of systemic infection	-0.77	<0.0001
α	Shedding rate in colon	0.62	<0.0001
β	Replication rate in colon	-0.60	<0.0001
μ	Migration rate to the caecal lymph node	-0.09	0.004
θ	Fraction of bacteria replicating in phagocyte	-0.09	0.005
C	Number of bacteria in phagocyte at rupture	-0.04	0.201

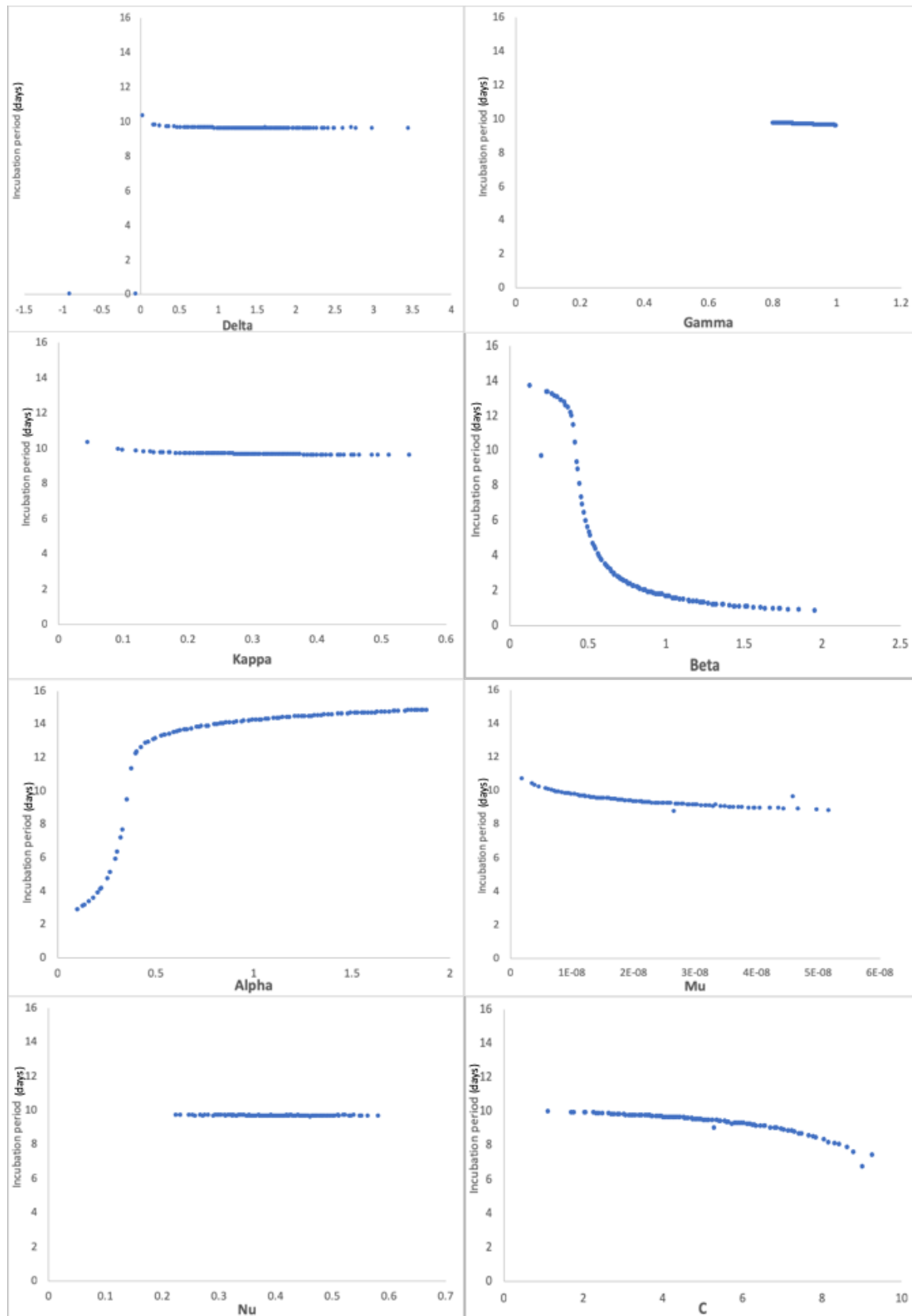


Figure 36 Scatterplots showing the relationship of each parameter with the incubation period (parameters: delta to C)

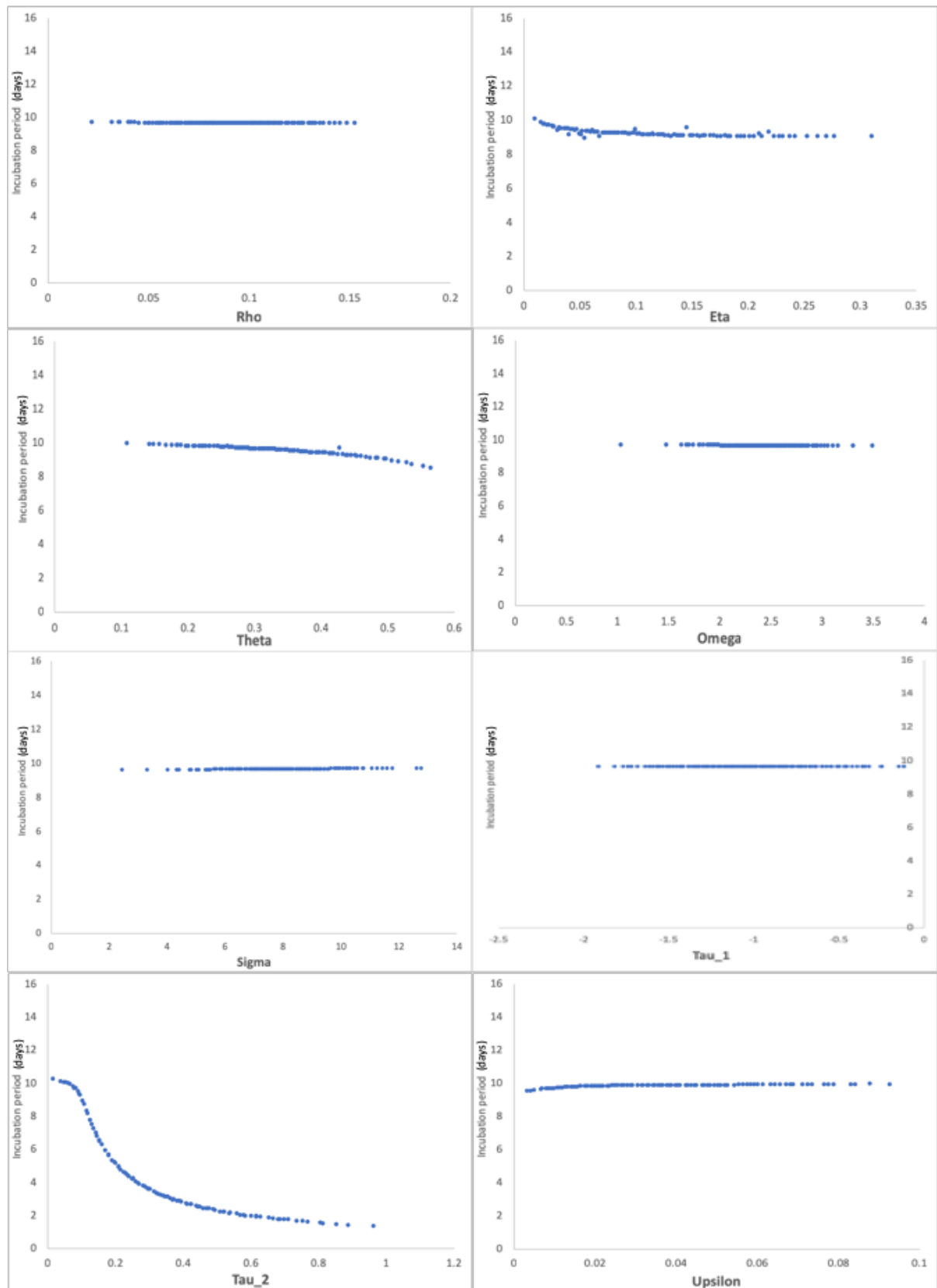


Figure 37 Scatterplots showing the relationship of each parameter with the incubation period (parameters: rho to upsilon)

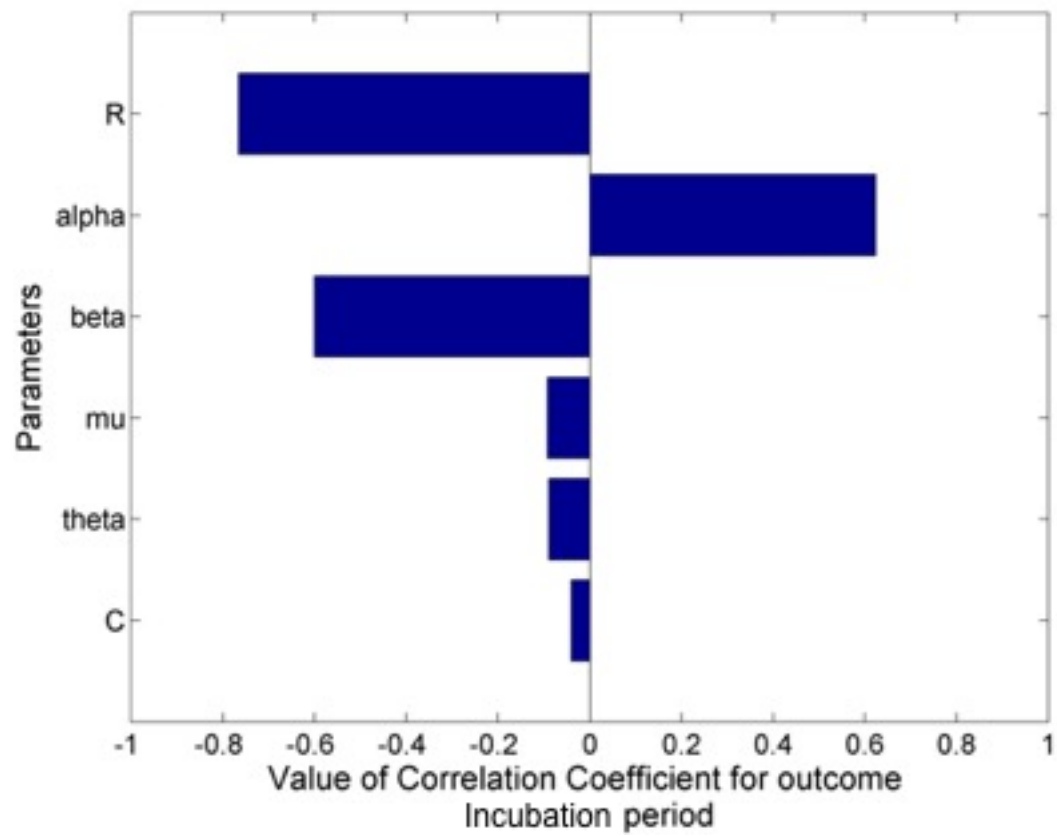


Figure 38 Tornado plot showing direction and strength of correlation with incubation period

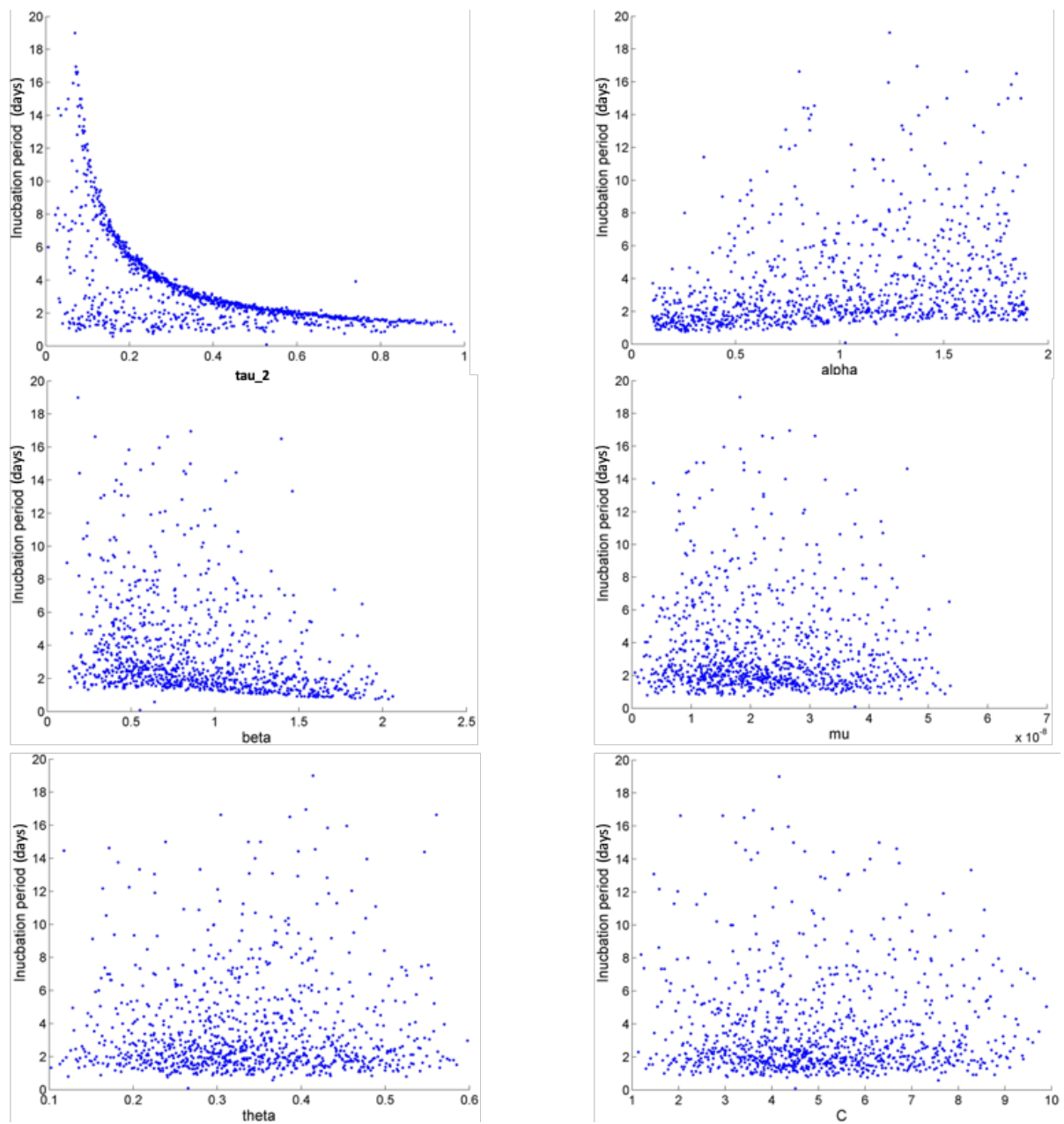


Figure 39 Scatterplots showing correlation of parameters with incubation period in the sensitivity analysis

9.8.3 Uncertainty on the predicted incubation period

Due to the uncertainty in the parameter values, the incubation period predicted by the model is expected to have a measure of uncertainty. The output of the sensitivity analysis and the simulated models show the range of likely incubation period (Figure 40). These simulated

values are determined by the values of τ_2 , α and β , and represent the variability in the incubation period.

The predicted incubation period from the model was 9.6 days and the plausible range of the predicated model (2.5% and 97.5% quintiles) was 0.9 days and 16.6 days.

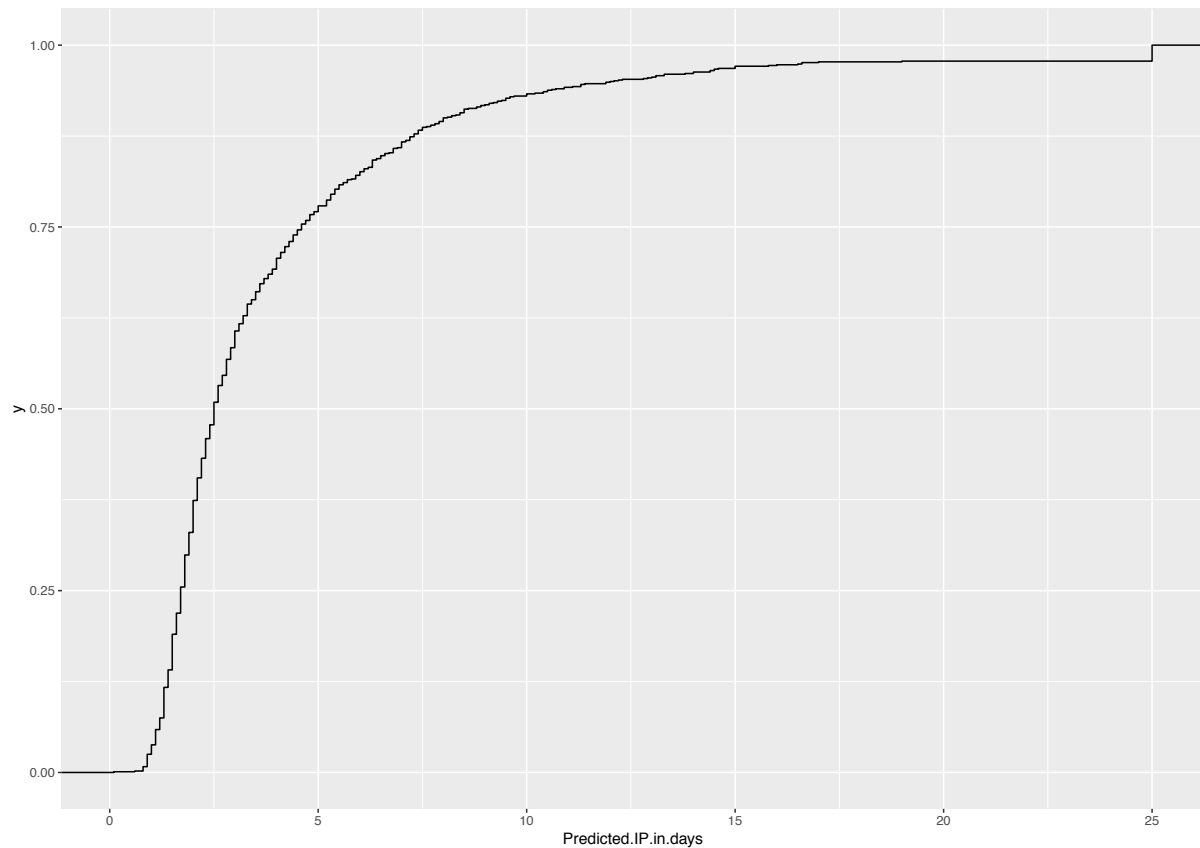


Figure 40 Empirical distribution function plot showing the predicted incubation period from 1000 simulated models

9.9 DISCUSSION

I have developed a compartmental model of ordinary differential equations to simulate the infection process of *S. Typhi* in humans in order to estimate the incubation period and identify biological processes that might influence the duration of incubation period. As far as I know, this is the first mathematical model on the incubation period of *S. Typhi* in humans.

Gastric acid acts as a barrier to most bacterial organisms including *S. Typhi* (260,261), restricting the proportion of bacteria that continue towards infection. However, its effect can be minimised due to the presence of any condition (drugs or disease) that reduces the stomach acid (261) or the constituents of the ingested food vehicle (129). Reduced gastric acid may result in shorter incubation period as more bacteria will survive to transit through to cause infection. In this model, the effect of gastric acid was eliminated, allowing nearly all bacteria to exit the stomach, mirroring the experimental study by Waddington et al (65) where sodium bicarbonate was used to restrict the effect of gastric acid.

The bacterial population in the colon is mainly determined by the rate of shedding (α) and the replication rate (β), hence, the difference of both parameters represents the net growth rate of bacterial population in the colon. If the net population of bacteria is greater than one ($\beta > \alpha$), infection would most likely proceed to the systemic phase as bacteria will be available in the colon to migrate into the CLn as observed in Figure 36. The interaction between α and β means that increasing the value of either parameter produces contrasting effects of equal magnitude on the incubation period as shown in the tornado plot (Figure 38).

The quantitative microbiology of *S. Typhi* in stool is currently unknown, hence, a biologically plausible value of $\beta = 1.8\alpha$ had to be assumed. This introduced some uncertainty into the model which was further examined in the sensitivity analysis. The output of this analysis indicated that α is an important determinant of the incubation period. The value of α is biologically plausible and was deliberately selected using the Waddington study as a validation dataset independent of the parameterisation.

According to the model, bacteria persists in the colon for the duration of the infection process (Figure 32) and this is consistent with the literature where shedding has been reported to occur for the duration of infection and even for a few days after antibiotic treatment (215). In cases without treatment, shedding can occur for up to three months (215,262). An experimental study showed that participants had positive stool cultures for as long as three weeks and all participants were culture negative 6 weeks after challenge (263). In another challenge study, stool cultures were positive from 72 hours after ingestion and for the two week duration of the experiment when samples were collected (65), however, the proportion of participants shedding (65) and probability of stool shedding (263) declined as the days post challenge increased.

Another critical point for infection, which also represented a major bottle neck, was the migration of bacteria to the CLn (μ). At this point, bacteria had a very low probability of migration and the likelihood of infection occurring largely depended on the bacteria crossing this threshold, thus encountering a major bottleneck. This was observed in a mouse experimental study conducted by Kaiser et al (227). According to the sensitivity analysis, an increase in the migration rate will shorten the incubation period as the process of systemic invasion is quickened.

Replication of bacteria in the liver and spleen during late phase systemic infection represents the last critical point for infection. Bacterial growth in this phase was strongly correlated with incubation period and from Figure 39, we can deduce that this is a dominant parameter for the incubation period distribution. As the growth rate increases, the number of bacteria available to enter into the bloodstream increases, and ultimately this reduces the incubation period. This conclusion is logical as it is the final opportunity for growth in the last step of the infection process before the onset of secondary bacteraemia. It is therefore plausible for any changes to bacterial growth rate at this point to significantly alter the incubation period.

The concentration of bacteria in the blood during secondary bacteraemia is predominantly made up of bacteria exiting the spleen and liver, however, according to the sensitivity analysis, the exit rate has limited effect on the duration of incubation period. Very early on in the onset of secondary bacteraemia, the spleen contributed more to the bacterial concentration, however, as infection progressed, the bacterial concentration was sustained by bacteria exiting the liver. This is shown in Figure 34 where the bacterial population in the liver was initially higher than the spleen, but subsequently, the bacterial population in the spleen increased and remained consistently higher for the duration of the infection. The identical replication rate (represented by τ) in both organs coupled with the difference in entry and exit rates, with the liver having higher rates, may explain the alteration in bacterial concentration and contribution observed between the liver and spleen.

Comparing the output of the model to data from a human experimental study, the predicted incubation period from the model was 9.6 days. This value was similar to estimates from other experimental and observational studies. In the systematic review conducted to estimate the distribution of incubation period and identify influencing factors (chapter 7), the mean incubation period between subgroups was reported to range from 9.7 days to 21.2 days with previously vaccinated cases reporting longer incubation period. A large experimental study conducted over 16 years (193), reported a mean incubation period of 11.4 days with a median of nine days. Another experimental study reported a shorter incubation period of a median of six days (IQR 5.1 to 7.8) in a control group with no vaccination who received a challenge dose of about 10^4 (264), similar to our initial dose. In an observational study of a large outbreak associated with Spanish spaghetti, the mean incubation period was 10.5 days with a median of eight days (187). A community outbreak associated with corned beef reported a mean incubation period of 9.1 days with a median of nine days (251).

According to the model, the bacterial levels in the blood at the time of diagnosis, as defined by Waddington et al (65), was 0.07 CFU/ml which was lower than the 1.1 CFU/ml they

reported, although similar values of low bacterial concentration of about 0.1 CFU/ml have been detected in the blood of typhoid cases (265,266).

The output of the model has shown that ingested dose, which could be a proxy for attack rate, influences the duration of the incubation period. In both the mathematical model and the experimental study, a ten-fold increase in the ingested dose reduced the incubation period by one day. This relationship has also been reported in some observational studies where an increase in the ingested dose shortened the incubation period (173,189).

Although the effect of vaccination has not been considered in this model, vaccination has been reported to prolong the incubation period (267). The effect of vaccination has been studied in an experimental study (264) where the time to microbiological diagnosis was similar in participants with and without vaccination, however, the median time to clinical diagnosis which also represented the incubation period was longer in the vaccine groups with reports of 8.5 and 10.4 days compared to the control group reporting 6.8 days. Some vaccines did not appear to affect the shedding pattern or attack rate (263,264), however, the level of bacteraemia was lowered. In order to achieve this effect, we can speculate that the most likely point of vaccine action will be restricting the replication of bacteria in the systemic organs, thereby reducing the rate of τ_2 .

According to Grant et al (77), individual immunological mechanisms can have different progressive effect on the bacterial population. Hence the immune response in one phase of the infection may be different from another phase of the infection as the bacteria encounters resident phagocytes, polymorphonuclear neutrophils and intracellular control mechanisms during the course of infection. The reduced exit rate of bacteria from the systemic organs to the blood at one percent of the initial entry rate may represent another possible adaptive immune response to slow down the process of infection. At this stage, the bacteria have survived all immunological mechanisms and the onset of sepsis is inevitable, however,

slowing down this process might enable further innate responses to intervene or maybe even allow acquired responses to kick in, intercepting the onset of secondary bacteraemia.

Our model was a deterministic model. Given the unknown shedding rate, we have forced bacterial growth in the colon by making $\beta > \alpha$ to ensure that infection progresses to the systemic phase. This however could result in unnatural timing given the fractional entry of bacteria into the CLn and so a stochastic model framework might have advantages.

Although this is a limitation of the model, it does not undermine the output or the knowledge contributed by this work as it is the first attempt at within host mathematical modelling of the incubation period of *S. Typhi* in humans. However, further work involving a stochastic model framework is recommended.

9.10 CONCLUSIONS

This chapter showed the application of mathematical models in understanding biological processes and predicting parameters of infectious diseases. The model has been useful in identifying factors intrinsic to the infection process that influence incubation period and suggests that the late phase net replication in organs and net replication rate in the colon are key determinants of the duration of the incubation period and may explain variation at a population level as much as ingested dose. Other factors such as vaccination have not been examined in this study and with the available evidence on the effect of vaccination, further work is required to extend this model to include vaccination.

CHAPTER 10: DISCUSSION

Knowledge of the incubation period, time between exposure to the infectious pathogen and the onset of clinical illness, is essential for several reasons which include: understanding the pathogenesis of an organism, investigating outbreaks and informing public health interventions and risk assessments.

This research project aimed to understand and describe the distribution of incubation period of some common gastrointestinal pathogens including *Campylobacter*, STEC and *Salmonella* Typhi and identify factors that may influence the distribution of incubation period. The methods used in this research included: systematic review and meta-regression, individual patient data analysis and mathematical modelling.

In synthesising data from observational studies, particularly data reporting measurement of parameters such as incubation period or dose response, traditional methods of meta-analysis of data and test for heterogeneity may not be sufficient. In addition to deriving the I^2 value and creating a forest plot, which made use of the mean of the distribution, methods employed in this research included KS test and hierarchical cluster analysis. The KS test compared the cumulative distributions of the reported data and was based on the distributional variation of the data rather than just the mean. Likewise, the hierarchical cluster analysis, using the output of the KS test as a measure of distance, described the pattern of heterogeneity based on the underlying distribution of data. The protocol describing the combination of traditional methods with new approaches was developed during the synthesis of data extracted from my review of *Campylobacter* and then applied in synthesising data from the reviews of STEC and *S. Typhi*.

The I^2 test and KS test indicated the presence of substantial heterogeneity in the reported incubation period data across all three reviews. Results from the I^2 tests, interpreted using the rough guide developed by Higgins et al (99), showed that the percentage of total variation across studies due to heterogeneity was high. Likewise, the resulting proportions of

pairwise comparisons in the KS tests showed evidence of differing distributions compared to the proportion expected by chance (5%) if the data were drawn from the same distribution. Furthermore, the re-created epidemic curves and the subgroup analyses showed that the heterogeneity observed is not only in relation to the mean incubation period, but also the shape of the distribution.

The variation observed in the incubation period between outbreaks could not be explained by chance alone, indicating that there may be underlying characteristics and factors responsible for the patterns seen. Some of the factors and characteristics that may explain the observed variation were identified by meta-regression and subgroup analyses of the hierarchical clusters. Across all three reviews, features qualitatively identified with subgroups mapped consistently with regression identified factors associated with incubation period. Food vehicle and age distribution were identified as characteristics associated with the incubation period of *Campylobacter* where outbreaks involving mostly children, and stating dairy product as the implicating food vehicle reported longer incubation periods. Factors associated with the incubation period of STEC were identified as age and attack rate. Children reported a shorter incubation period and an increase in the attack rate was associated with a reduction in the incubation period. Similarly, an increase in the attack rate resulted in a reduction in the incubation period of outbreaks associated with *S. Typhi*. Vaccine history was also identified as an influencing factor as previously vaccinated cases reported longer incubation period.

The type of data and amount of information available from outbreak reports made it difficult to examine additional patient characteristics that may influence the incubation period. Analysing individual patient data of cases associated with STEC outbreaks provided the opportunity to examine some patient characteristics such as individual age and reported symptoms. Characteristics associated with the outbreaks, such as mode of transmission, could also be examined. The mean incubation period of cases in the individual patient data analysis was similar to what was observed in the systematic review. Likewise, I observed the

presence of heterogeneity in the reported incubation period across cases and also across outbreaks which could not be explained by chance alone. Bloody diarrhoea and mode of transmission were identified as factors influencing the incubation period distribution with bloody diarrhoea being associated with shorter incubation period and outbreaks associated with direct animal contact and foodborne exposure reporting longer incubation periods. The individual patient data was a collation of information on individual cases, hence, detailed information on the outbreaks and the exposed population, which were available in outbreak reports, were scarce and inaccessible. Thus, the attack rates of the outbreaks were unavailable for analysis.

The systematic reviews and analysis of individual patient data examined outbreak characteristics and patient characteristics that may explain heterogeneity; therefore, the within host mathematical model was developed to examine how the steps involved in the process of infection contribute to the incubation period and how variation and uncertainty at each step contributes to overall uncertainty and variation in the incubation period.

Characteristics like age and food vehicle were not examined here as they were assumed to be constant. The process of developing the model required in-depth knowledge of the fundamental concept, hence, a literature review on the pathophysiology of *S. Typhi* was first conducted.

The model was parameterised using both human and mouse experimental studies, and the predicted incubation period was 9.6 days, similar to the output of the systematic review. The reliability of the model output largely depended on the model parameterisation and the reliability of the parameter values. Selecting these values involved reviewing a number of experimental studies, and sometimes, different values were identified for the same parameter. The decision to select one value over another was made in a systematic way based on a hierarchy of evidence (Figure 28) and the context of the supporting evidence.

DISTRIBUTION OF INCUBATION PERIOD DATA

This study confirmed that the distribution of incubation period data may not necessarily follow a lognormal distribution as previously assumed in literature (268). Due to the obvious skew inherent in incubation period data other distribution that may be considered include the gamma distribution and the generalised beta distribution, both special cases of the Pearson family of distributions (269). The generalised beta is also called the Pearson Type I distribution and gamma distribution is called the Pearson Type III.

Given the combined incubation period distribution of *Campylobacter* data from all studies, the Cullen and Frey graph (270) showed the data to have consistent skew and kurtosis with a generalised beta distribution but when bootstrapped to allow for data uncertainty the gamma distribution is a plausible and more parsimonious model. The data were then fitted to a gamma distribution by maximum likelihood, and the goodness of fit plots indicated that the incubation period data was well characterised by a gamma distribution.

For both *S. Typhi* and STEC data, the observed and bootstrapped data points were distributed around the generalised beta region of the Cullen and Frey graphs, suggesting other families of distributions unsuitable (not sufficiently explaining the skew and kurtosis of the data). The Pearson Type I distribution is restricted to upper and lower limits, which may not be biologically plausible so caution is potentially necessary in interpretation. Fitting a Pearson Type 1 distribution to the data using method of moments (271) we obtain the parameter values (shape 1 and shape 2) and the values for the lower and upper limit (Table 27). This shows that the upper limit in two cases is relatively large (given the largest observed period in the actual data) and so might indicate overfitting to the data. Further investigation is necessary to determine the justification of such families of distributions to incubation period data.

Table 27 Parameters obtained fitting Pearson Type 1 distribution to the incubation period data

Data	Shape 1	Shape 2	Lower limit	Upper limit
Campylobacter	0.18	22.78	1.32	485.79
STEC	1.99e-02	1.58e+00	5.45e+00	3.97e+03
S. Typhi	0.15	1.90	1.22	23.79

STRENGTHS AND WEAKNESSES OF METHODS AND DATA

Systematic review

Using a systematic review to identify reported incubation period in the literature made it possible to combine several outbreaks or experimental studies investigating the infecting pathogen, likewise, conducting a meta-regression on the combined extracted data meant that reliable and accurate conclusions can be drawn (272).

The outbreaks and experimental studies included in the reviews were not primarily designed to study the incubation period of the pathogen but reporting the outcome of an outbreak investigation or studying other aspects of the pathogens, hence, the reported data could not be assessed for quality as in a traditional review (68). This however meant that the observed incubation period did not influence the decision to publish, thereby reducing the likelihood of publication bias and selection bias.

The reported data were available in different formats with studies reporting several units of incubation times e.g. hourly, six-hourly, days and aggregated two-days. Some extracted data had to be converted to hours or days and were either rounded up or rounded down which could have resulted in an overestimation and underestimation of the incubation period.

The information provided in the included studies were usually a summary of the study population and thus not detailed enough to evaluate other possible factors influencing incubation period such as underlying conditions, current medications and previous infections. Outbreak reports cannot contain the same detail of data available in individual patient data, therefore where possible, data repositories of outbreak data and anonymised individual patient data should be created such that they can be jointly analysed and contribute to the wider evidence base. Such repository would collate study data, protocols, data collection tools and other documents from foodborne and non-foodborne outbreaks.

Analysis of Individual patient data

Individual patient data analysis can be conducted as part of a systematic review, but in this research, it has been conducted as a standalone analysis; collating all cases of STEC that have been reported as part of an outbreak in England and Wales. This type of data results in harmonised analysis and outcomes (273) as well as comparison across outbreaks.

Analysing the incubation period across different patient types and characteristics is made possible through individual patient data as further exploration is possible beyond summary data.

The individual patient data used in this research was collated as part of an enhanced surveillance system. This is an active surveillance system that aims to identify every case of STEC, either sporadic or outbreak related. Using data from a surveillance system ensures a wide coverage of the study population and the generalizability of the results to the overall population. As part of the surveillance activity, information is collected from every identified case. The information collected includes demographic details, presenting symptoms, exposure details and treatment information. This provides a wealth of information on each case that can be analysed and studied with reference to the incubation period.

However, this type of data also has its limitations stemming from restricted access to the data, the incompleteness of the collected data and unavailability of data. The information collected by the surveillance system includes patient identifiable data which is protected by law. Using such information requires the user to be granted permission to access the data by the owners, in this case, Public Health England. Furthermore, the data will have to be de-identified and may only be accessible under certain conditions in line with UK data policy and guidelines (274). The process of accessing the data could be prolonged, thereby delaying the onset of research. In the course of this research project, I sought for permission and access to data from the enhanced surveillance system for two years, and when this was finally granted at the end of my second year, the data could only be viewed and analysed at a Public Health England site, using an encrypted computer.

Although the data collected by the surveillance system is supposed to be comprehensive, the structure of the questions in the standardised questionnaire does not capture specific exposure time. A large proportion of the available records, particularly cases associated with foodborne outbreaks, did not include information on the date of exposure making it impossible to calculate the incubation period of cases. This was predominantly seen in non-point source outbreaks where the case might remember eating a particular food in a time period, but not the actual day or time of consumption. As is the case in an enhanced surveillance system, several cases would have been identified and the questionnaire completed before being classified as part of an outbreak, thus, some relevant information may be overlooked at the time of interview. In this research, the characteristics of the cases with missing data were significantly different from those with complete data, therefore, the possibility of bias cannot be ruled out. Furthermore, for enhanced surveillance data to be used for any type of research, it must first be available. Retrospectively collecting data increases the risk of historic records being lost or missing for any reason.

As the surveillance system was not developed for the purpose of studying the incubation period of STEC, certain information that would have been relevant in understanding its

distribution were not available, such as the onset of HUS, underlying medical conditions, ongoing medications or previous STEC infections.

Mathematical modelling

In natural experiments or outbreak situations, the consistency and contents of ingested food at the time of exposure may alter the duration of gastric emptying or the effect of gastric acid, which in turn affects the incubation period (129,219). However, in the mathematical model, I assumed that the food vehicle is liquid that also protects the bacteria from the effect of gastric acid, such as sodium bicarbonate or milk, which may not be representative of real-life situations.

As with most models, data gaps are identified during model formulation and parameterisation. I was unable to find any evidence supporting the quantitative microbiology of *S. Typhi* in stool. The value assumed for this parameter was based on biological plausibility; nevertheless, values in observational studies may differ. Results from the sensitivity analysis conducted indicated that bacterial shedding from the colon is strongly correlated with the incubation period showing the importance of this evidence gap. As a result of the unknown variation in shedding rate, the actual effect of bacterial shedding on incubation period could not be examined.

The dynamics of a model can be studied as parameters can be excluded or changed to see the effect on the model output. It can also be easily manipulated for experimentation and prediction. While conducting the sensitivity analysis, I ran the model with simulated values randomly selected from the parameter distribution to identify parameters that were mostly associated with the incubation period. Using mathematical models also provides a cheaper and quicker alternative to study infectious diseases, reducing the time it takes to arrive at a conclusion.

Another advantage with using mathematical models is the opportunity to study certain aspects of the infection process and how they contribute to the incubation period; that would otherwise be difficult to study from observational or experimental studies. One of such aspects is the infection of *S. Typhi* at the cellular level involving lymph nodes and phagocytes.

PRACTICAL IMPLICATIONS OF STUDY FINDINGS

A consistent finding observed from the systematic reviews and analysis of individual patient data was the heterogeneity in the reported and estimated incubation period between outbreaks and between individuals. Some of the possible reasons identified for the observed heterogeneity were patient's age, patient's medical history including vaccine history, presenting symptoms, and mode of transmission.

As these factors have varying effects on the incubation period, public health experts should be aware that a uniform standard approach will not be efficient in reducing the burden of disease brought on by gastrointestinal infections. Therefore, outbreak investigations, risk assessments and public health response should be designed to accommodate the underlying variation of the population at risk and the outbreak characteristics.

Characteristics of the affected population and the infecting pathogen should inform the process of outbreak investigations. For instance, trawling questionnaires completed during an outbreak of campylobacteriosis involving children alone should collect exposure data over a longer time period in order to identify the potential source of infection as longer incubation periods can be expected. The evidence available on the effect of STEC serogroup on incubation period was dominated by O104. The lack of other non-O157 serogroups made it difficult to confirm if the resulting incubation periods are longer or shorter. Nevertheless, the non-O157 outbreaks (O104) identified reported longer incubation periods, thus, the process of investigating outbreaks of STEC O157 and non-STEC O157 should not be identical, considering the potential difference in their incubation times.

In developing standardised questionnaires for the purpose of surveillance, outbreak control or research, specific questions that may inform the expected incubation period should be included. Such questions include the vaccine history for cases of typhoidal salmonellosis or the presence and onset of bloody diarrhoea in STEC cases. Other information which may be

useful to collect in order to examine the relationship with incubation period include co-morbidities, ongoing medication and the development of HUS in STEC cases.

The attack rate of an outbreak was observed to influence the incubation period in the systematic review of both *Salmonella* Typhi and STEC, perhaps because a high attack rate is associated with higher doses of pathogen from a heavily contaminated source which could also affect incubation period. Where attack rate can be calculated, particularly in a point source outbreak, this information may be used as an early indication of the expected incubation period and subsequent outbreak investigations can be adapted accordingly.

In addition to patient and outbreak characteristics, public health experts also need to be aware of the variation in host physiology which may influence the incubation period of a disease. According to the mathematical model, net growth of bacteria in the colon and the transfer rate of bacteria into the blood (for systemic infections) were identified as factors that influence the incubation period. Gastric emptying rate, which has been reported to differ by gender and food type (liquid or solid), may lengthen or shorten the incubation period accordingly. Similarly, the effect of gastric acid, which can be altered by medication, physiological conditions or even the ingested food, can also affect the resulting incubation period. These factors and other physiological processes which are not assessable during infection or in an outbreak situation should be considered when investigating outbreaks.

FURTHER RESEARCH RECOMMENDED

In the absence of standard guidelines in identifying statistically reasonable cut-off points in hierarchical cluster analysis, I selected a value which may not be generalizable to other research. Further work is needed to rationalise choice when applied to other datasets and reviews.

The development of a data repository combining outbreak data and individual patient data is recommended for Public Health England to support research. This will be a shared resource of archived outbreak investigations including reports, data collection instruments, and datasets to support PHE in fully using data collected to contribute to health protection by analysis across outbreaks and to quality improvement in outbreak investigation.

There was limited evidence on the quantitative microbiology of *S. Typhi* in stool and also the flow rate of the caecal lymph node into the intestinal lymph vessels. The values selected during model parameterisation were biologically plausible best guess estimates or reports without supporting experimental studies. Further research is needed to study these biological processes and how it contributes to the course of infection. Furthermore, developing a stochastic model framework on incubation period is suggested to account for the very low rate of migration to the CLn.

PERSONAL REFLECTION

Completing this research project involved a steep learning curve technically, professionally and personally some of which were self-taught where I learnt how to use the R software for statistical analysis and how to conduct sensitivity analysis using probability distributions with SaSAT software.

Undertaking the systematic reviews and conducting a literature review for the model parameterisation greatly improved my understanding of the subject matter. The process also developed my critical appraisal and bibliography skills. As part of the study selection process in the systematic review, I was able to work with other students and researchers which helped me work better as part of a team.

In order to analyse individual patient data, I had to request for data from Public Health England (PHE). This process was longer than expected as stages of approval were required. At the beginning of my DPhil program, my plan was to collate outbreak information on foodborne and non-foodborne diseases across the various health protection unit of PHE and create a database or archive of outbreak investigation resources. This archive was to include study data, protocols, data collection tools and other document involved in the investigation of outbreaks. Once this archive was completed, the individual patient data of cases involved in outbreaks of *S. Typhi*, *Campylobacter*, *Clostridium perfringens* and STEC were to be extracted for analysis. Unfortunately, the PHE approval process took two years, and by the time this was completed, it was no longer practical to create the outbreak archive as originally planned within the time I had left with my funding body. Hence, the STEC enhanced surveillance database was used as an alternative as it already collated individual patient information on cases of STEC in England and Wales. My interaction with PHE improved my negotiation skills and I learnt how to work through bureaucratic processes while ensuring that I achieved my aims.

At the start of my program, I had very little knowledge of mathematical modelling. During the course of my project, I attended two training courses on the mathematical modelling of infectious diseases offered by Imperial College and the Wellcome Trust. I acquired the technical expertise to develop ordinary differential equations and solve models using the Berkeley Madonna software and R statistical package. The courses also provided a good avenue to network and develop professional relationships with other students and instructors which has proven to be useful.

During the course of my program, I have had the opportunity of presenting my findings at international and national conferences as well as scientific group meetings. One of such conferences was the 2018 International Conference on Emerging Infectious Diseases (ICEID) in Atlanta where I presented my work on the incubation period of *S. Typhi*. Submitting abstracts and preparing materials for oral and poster presentations has enhanced my presentation and communication skills.

Overall, my DPhil program has been as challenging as it has been rewarding, developing my ability to independently conduct a research project. Most of the research project involved me working on my own with support from my supervisors. I was able to manage my time efficiently, setting deadlines based on my timetable, and producing outputs and reports in good time for review by my supervisors. This made me self-motivated and resilient, with the ability to work well on my own and also with other people. My data analysis and management skills has been greatly improved with the knowledge of how to use three software packages (R statistical software, Berkeley Madonna and SaSAT). So far, the output of my research has resulted in three peer reviewed publications, with three further manuscripts undergoing review process. Writing these papers as well as my thesis has helped me fine-tune my report writing skills, enabling me to present the necessary information in a concise way while concentrating on the facts.

The support I received from my supervisors has been excellent all through my program. Prof Noel McCarthy and Dr Ian Hall were always available to review my work, brainstorm on possible analytical methods and provide guidance and direction as the research progressed. For research areas I was unfamiliar with at the beginning of the research project, such as the mathematical modelling, they were willing and able to teach and explain things explicitly for me to understand. Furthermore, under their supervision, I have become an independent researcher and an expert in infectious disease epidemiology with additional technical skills.

The NIHR Health Protection Research Unit in GI infections, which is my funding body provided an atmosphere where similarly funded students could come together under formal and informal settings. This was a good avenue for networking, but I also found it as a way of having a sense of belonging to a wider community.

In light of the expertise, skills and knowledge I have gained during my research program, life after DPhil looks very promising. I am looking forward to working with international health organisations or pharmaceutical companies, where I can apply what I have learnt into real life situations and also further develop myself professionally.

REFERENCES

1. LaRocque RC, Calderwood SB. 98 - Syndromes of Enteric Infection. In: Bennett JE, Dolin R, Blaser MJ, editors. *Mandell, Douglas, and Bennett's Principles and practice of infectious diseases (Eighth Edition)* [Internet] [cited 2018 Sep 24]. p. 1238-1247.e2. Available from: <http://www.sciencedirect.com/science/article/pii/B9781455748013000989>
2. Fletcher SM, McLaws M-L, Ellis JT. Prevalence of gastrointestinal pathogens in developed and developing countries: systematic review and meta-analysis. *J Public Health Res.* 2013 Jul 16;2(1):42–53.
3. Esan OB, Pearce M, van Hecke O, Roberts N, Collins DRJ, Violato M, et al. Factors associated with sequelae of campylobacter and non-typhoidal salmonella infections: a systematic review. *EBioMedicine.* 2017 Feb 1;15:100–11.
4. Musher DM, Musher BL. Contagious acute gastrointestinal infections. *N Engl J Med.* 2004 Dec 2;351(23):2417–27.
5. Riddle MS, DuPont HL, Connor BA. ACG Clinical Guideline: diagnosis, treatment, and prevention of acute diarrheal infections in adults. *Am J Gastroenterol.* 2016;111(5):602–22.
6. World Health Organization. Foodborne disease outbreaks: Guidelines for investigation and control [Internet]. 2008 [cited 2016 Nov 17]. Available from: http://www.who.int/foodsafety/publications/foodborne_disease/outbreak_guidelines.pdf
7. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Lond Engl.* 2016 08;388(10053):1545–602.
8. GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Lond Engl.* 2016 Oct 8;388(10053):1459–544.
9. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLOS Med.* 2015 Dec 3;12(12):e1001921.
10. Food Standards Agency. The second study of infectious intestinal disease in the community (IID2 Study) | Food Standards Agency [Internet]. [cited 2016 Nov 9]. Available from: <https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021>
11. Roberts JA, Cumberland P, Sockett PN, Wheeler J, Rodrigues LC, Sethi D, et al. The study of infectious intestinal disease in England: socio-economic impact. *Epidemiol Infect.* 2003 Feb;130(1):1–11.

12. Thomas M, Majowicz SE, Pollari F, Sockett P. Burden of acute gastrointestinal illness in Canada, 1999-2007: Interim summary of NSAGI activities. *Can Commun Dis Rep*. 2008 May;35(5).
13. Ternhag A, Törner A, Svensson Å, Ekdahl K, Giesecke J. Short- and long-term effects of bacterial gastrointestinal infections. *Emerg Infect Dis*. 14(1):143–8.
14. World Health Organization. The global view of campylobacteriosis: Report of an expert consultation. [Internet]. Utrecht, Netherlands; 2012 Jul. Available from: http://apps.who.int/iris/bitstream/handle/10665/80751/9789241564601_eng.pdf;jsessionid=2009568E665C4D38FA30424761546E0F?sequence=1
15. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of campylobacter infection. *Clin Microbiol Rev*. 2015 Jul 1;28(3):687–720.
16. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*. 2011 Jun 27;gut.2011.238386.
17. Hall G, Yohannes K, Raupach J, Becker N, Kirk M. Estimating community incidence of *Salmonella*, *Campylobacter*, and Shiga Toxin–producing *Escherichia coli* infections, Australia. *Emerg Infect Dis*. 2008 Oct;14(10):1601–9.
18. Ruiz-Palacios GM. The Health burden of *Campylobacter* infection and the impact of antimicrobial resistance: Playing Chicken. *Clin Infect Dis*. 2007 Mar 1;44(5):701–3.
19. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid fever. *Clin Infect Dis*. 2010 Jan 15;50(2):241–6.
20. WHO | Typhoid [Internet]. WHO. [cited 2017 Oct 2]. Available from: <http://www.who.int/immunization/topics/typhoid/en/>
21. Crump JA, Luby SP, Mintz ED. The global burden of typhoid fever. *Bull World Health Organ*. 2004 May;82(5):346–53.
22. Public Health England. Enteric fever (typhoid and paratyphoid) in England, Wales and Northern Ireland: 2016 [Internet]. 2017 Nov [cited 2018 Apr 16]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/657154/enteric_fever_annual_2016_report.pdf
23. Antillón M, Warren JL, Crawford FW, Weinberger DM, Kürüm E, Pak GD, et al. The burden of typhoid fever in low- and middle-income countries: A meta-regression approach. *PLoS Negl Trop Dis*. 2017 Feb 27;11(2):e0005376.
24. Haagsma JA, Geenen PL, Ethelberg S, Fetsch A, Hansdotter F, Jansen A, et al. Community incidence of pathogen-specific gastroenteritis: reconstructing the surveillance pyramid for seven pathogens in seven European Union member states. *Epidemiol Amp Infect*. 2013 Aug;141(8):1625–39.
25. Kaper JB, O'Brien AD. Overview and historical perspectives. *Microbiol Spectr* [Internet]. 2014 [cited 2018 Apr 12];2(2). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4290666/>
26. Majowicz SE, Scallan E, Jones-Bitton A, Sargeant JM, Stapleton J, Angulo FJ, et al. Global incidence of human Shiga Toxin–Producing *Escherichia coli* infections and

- deaths: a systematic review and knowledge synthesis. *Foodborne Pathog Dis.* 2014 Jun;11(6):447–55.
27. Painter JA, Ayers T, Woodruff R, Blanton E, Perez N, Hoekstra RM, et al. Recipes for foodborne outbreaks: a scheme for categorizing and grouping implicated foods. *Foodborne Pathog Dis.* 2009 Dec;6(10):1259–64.
 28. Painter JA, Hoekstra RM, Ayers T, Tauxe RV, Braden CR, Angulo FJ, et al. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008. *Emerg Infect Dis.* 2013 Mar;19(3):407–15.
 29. Adak GK, Meakins SM, Yip H, Lopman BA, O'Brien SJ. Disease risks from foods, England and Wales, 1996–2000. *Emerg Infect Dis.* 2005 Mar;11(3):365–72.
 30. Astridge K, McPherson M, Kirk M, Knope K, Gregory J, Kardamandis K, et al. Foodborne disease outbreaks in Australia 2001 - 2009. *ANU Res Publ [Internet].* 2011; Available from: <https://openresearch-repository.anu.edu.au/bitstream/1885/53247/8/Foodborne%20Disease%20Outbreaks%20in%20Australia%202001-2009%20-%20Food%20Australia%20%282011%29%20V.64%20iss.12%20pp44-50.pdf>
 31. Public Health England. Foodborne and non-foodborne gastrointestinal outbreaks surveillance [Internet]. [cited 2018 Oct 24]. Available from: <https://www.gov.uk/guidance/foodborne-and-non-foodborne-gastrointestinal-outbreaks-surveillance>
 32. Friedman CR, Hoekstra RM, Samuel M, Marcus R, Bender J, Shiferaw B, et al. Risk factors for sporadic *Campylobacter* infection in the United States: A case-control study in FoodNet sites. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 2004 Apr 15;38 Suppl 3:S285-296.
 33. Gallay A, Valk HD, Cournot M, Ladeuil B, Hemery C, Castor C, et al. A large multi-pathogen waterborne community outbreak linked to faecal contamination of a groundwater system, France, 2000. *Clin Microbiol Infect.* 2006 Jun 1;12(6):561–70.
 34. Lindmark H, Boqvist S, Ljungström M, Ågren P, Björkholm B, Engstrand L. Risk factors for campylobacteriosis: an epidemiological surveillance study of patients and retail poultry. *J Clin Microbiol.* 2009 Aug 1;47(8):2616–9.
 35. Taylor EV, Herman KM, Ailes EC, Fitzgerald C, Yoder JS, Mahon BE, et al. Common source outbreaks of *Campylobacter* infection in the USA, 1997-2008. *Epidemiol Infect.* 2013 May;141(5):987–96.
 36. Potter RC, Kaneene JB, Hall WN. Risk factors for sporadic *Campylobacter jejuni* infections in rural michigan: a prospective case-control study. *Am J Public Health.* 2003 Dec;93(12):2118.
 37. Ravel A, Pintar K, Nesbitt A, Pollari F. Non-food-related risk factors of campylobacteriosis in Canada: a matched case-control study. *BMC Public Health.* 2016 Sep 27;16:1016.
 38. Domingues AR, Pires SM, Halasa T, Hald T. Source attribution of human salmonellosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiol Amp Infect.* 2012 Jun;140(6):959–69.

39. Hosoglu S, Celen MK, Geyik MF, Akalin S, Ayaz C, Acemoglu H, et al. Risk factors for typhoid fever among adult patients in Diyarbakir, Turkey. *Epidemiol Infect.* 2006 Jun;134(3):612–6.
40. Vollaard AM, Ali S, van Asten Hagh, Widjaja S, Visser LG, Surjadi C, et al. Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. *JAMA.* 2004 Jun 2;291(21):2607–15.
41. Bryan FL. Risks of practices, procedures and processes that lead to outbreaks of foodborne diseases. *J Food Prot.* 1988 Aug 1;51(8):663–73.
42. Dauer C. Food and water borne disease outbreaks. *Public Health Rep.* 1952 Nov;67(11):1089–95.
43. Rivas M, Sosa-Estani S, Rangel J, Caletti MG, Vallés P, Roldán CD, et al. Risk factors for sporadic Shiga Toxin–producing *Escherichia coli* infections in children, Argentina. *Emerg Infect Dis.* 2008 May;14(5):763–71.
44. Rivas M, Chinen I, Miliwebsky E, Masana M. Risk Factors for Shiga Toxin-Producing *Escherichia coli*-associated human diseases. *Microbiol Spectr* [Internet]. 2014 Sep 12 [cited 2018 Apr 18];2(5). Available from: <http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.EHEC-0002-2013>
45. McPherson M, Lalor K, Combs B, Raupach J, Stafford R, Kirk MD. Serogroup-specific risk factors for Shiga Toxin-Producing *Escherichia coli* infection in Australia. *Clin Infect Dis.* 2009 Jul 15;49(2):249–56.
46. Olsen SJ, MacKinnon L, Goulding JS, Bean NH, Slutsker L. Surveillance for foodborne disease outbreaks --United States, 1993-1997. 2000 Mar 17;49(SS01):1–51.
47. Public Health England. Communicable disease outbreak management: operational guidance [Internet]. [cited 2018 Oct 24]. Available from: <https://www.gov.uk/government/publications/communicable-disease-outbreak-management-operational-guidance>
48. Nishiura H. Early efforts in modelling the incubation period of infectious diseases with an acute course of illness. *Emerg Themes Epidemiol.* 2007 May 11;4:2.
49. Sartwell P. Incubation period and the dynamics of infectious disease. *Am J Epidemiol.* 1986;83(2):204–16.
50. Mandell GL, Douglas RG, Bennett JE, Dolin R. Principles and practice of infectious diseases: antimicrobial therapy, 1996/1997. *Princ Pract Infect Dis Antimicrob Ther* 1996/1997 [Internet]. 1996 [cited 2018 Apr 19]; Available from: <https://www.cabdirect.org/cabdirect/abstract/19962006309>
51. Guzman-Herrador B, Vold L, Nygard K. Surveillance of travel-associated gastrointestinal infections in Norway, 2009-2010: are they all actually imported? *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull.* 2012;17(41):20294.
52. Horn BJ, Lake RJ. Incubation period for campylobacteriosis and its importance in the estimation of incidence related to travel. *Euro Surveill.* 2013;18(40).

53. Dhar J, Sharma A. The role of the incubation period in a disease model. *Appl Math E-Notes*. 2009;9:146–53.
54. Nishiura H, Mizumoto K, Ejima K, Zhong Y, Cowling BJ, Omori R. Incubation period as part of the case definition of severe respiratory illness caused by a novel coronavirus. *Euro Surveill Eur Commun Dis Bull [Internet]*. 2012 Oct 18 [cited 2016 Dec 21];17(42). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3722063/>
55. Brookmeyer R, You X. A hypothesis test for the end of a common source outbreak. *Biometrics*. 2006 Mar;62(1):61–5.
56. Review of the major outbreak of *E. coli* O157 in Surrey, 2009. - [Internet]. [cited 2015 Oct 26]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/342361/Review_of_major_outbreak_of_e_coli_o157_in_surrey_2009.pdf
57. Robert Koch Institute. Report: Final presentation and evaluation of epidemiological findings in the EHEC O104:H4 outbreak, Germany 2011. Berlin 2011;
58. Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. Experimental *Campylobacter jejuni* infection in humans. *J Infect Dis*. 1988 Mar 1;157(3):472–9.
59. Tribble DR, Baqar S, Scott DA, Oplinger ML, Trespalacios F, Rollins D, et al. Assessment of the duration of protection in *Campylobacter jejuni* experimental infection in humans. *Infect Immun*. 2010 Apr;78(4):1750–9.
60. Kuusi M, Nuorti JP, Hänninen ML, Koskela M, Jussila V, Kela E, et al. A large outbreak of campylobacteriosis associated with a municipal water supply in Finland. *Epidemiol Infect*. 2005 Aug;133(4):593–601.
61. Porter IA, Reid TM. A milk-borne outbreak of *Campylobacter* infection. *J Hyg (Lond)*. 1980 Jun;84(3):415.
62. Wood RC, MacDonald KL, Osterholm MT. *Campylobacter* enteritis outbreaks associated with drinking raw milk during youth activities. A 10-year review of outbreaks in the United States. *JAMA*. 1992 Dec 9;268(22):3228–30.
63. World Health Organization. Background document: The diagnosis, treatment and prevention of typhoid fever [Internet]. [cited 2017 Oct 10]. Report No.: WHO/V&B/03.07. Available from: <http://www.who.int/rpc/TFGuideWHO.pdf>
64. Waddington CS, Darton TC, Woodward WE, Angus B, Levine MM, Pollard AJ. Advancing the management and control of typhoid fever: A review of the historical role of human challenge studies. *J Infect*. 2014 May;68(5):405–18.
65. Waddington CS, Darton TC, Jones C, Haworth K, Peters A, John T, et al. An outpatient, ambulant-design, controlled human infection model using escalating doses of *Salmonella typhi* challenge delivered in sodium bicarbonate solution. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2014 May 1;58(9):1230–40.
66. Miner JR. The Incubation Period of Typhoid Fever. *J Infect Dis*. 1922;31(3):296–301.
67. European Centre for Disease Prevention and Control. Systematic review on the incubation and infectiousness/shedding period of communicable diseases in children [Internet]. Stockholm: European Centre for Disease Prevention and Control; 2016

[cited 2016 Oct 28]. Available from:
<http://ecdc.europa.eu/en/publications/publications/systematic-review-incubation-period-shedding-children.pdf>

68. Todkill D, Fowler T, Hawker JI. Estimating the incubation period of acute Q fever, a systematic review. *Epidemiol Infect.* 2018 Apr;146(6):665–72.
69. Goulet V, King LA, Vaillant V, de Valk H. What is the incubation period for listeriosis? *BMC Infect Dis.* 2013;13:11.
70. Chan M, Johansson MA. The Incubation Periods of Dengue Viruses. *PLOS ONE.* 2012 Nov 30;7(11):e50972.
71. Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating incubation period distributions with coarse data. *Stat Med.* 2009 Sep 30;28(22):2769–84.
72. Vynnycky E, White R. *An Introduction to Infectious Disease Modelling.* OUP Oxford; 2010. 402 p.
73. Neumaier A. Mathematical Modeling, Chapter 3. In: *Modeling Languages in Mathematical Optimization* [Internet]. Kluwer, Boston; 2004 [cited 2018 Sep 24]. (Applied Optimization; vol. 88). Available from: <https://www.mat.univie.ac.at/~neum/model.html>
74. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, Angelis DD, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science.* 2015 Mar 13;347(6227):aaa4339.
75. Fenton A. Editorial: Mathematical modelling of infectious diseases. *Parasitology.* 2016 Jun;143(7):801–4.
76. Keeling MJ, Danon L. Mathematical modelling of infectious diseases. *Br Med Bull.* 2009;92:33–42.
77. Grant AJ, Restif O, McKinley TJ, Sheppard M, Maskell DJ, Mastroeni P. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. *PLOS Biol.* 2008 Apr 8;6(4):e74.
78. Marchuk GI, Asachenkov AL, Belykh LN, Zuev SM. Mathematical modelling of infectious diseases. In: Hoffmann GW, Hraba T, editors. *Immunology and Epidemiology.* Springer Berlin Heidelberg; 1986. p. 64–81.
79. Kao RR. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends Microbiol.* 2002 Jun 1;10(6):279–86.
80. Pongsumpun P. Mathematical model of Dengue disease with the incubation period of virus. *Int J Math Comput Sci.* 2008;2(8):535–9.
81. Lui KJ, Lawrence DN, Morgan WM, Peterman TA, Haverkos HW, Bregman DJ. A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome. *Proc Natl Acad Sci.* 1986 May 1;83(10):3051–5.
82. Brookmeyer R, Johnson E, Barry S. Modelling the incubation period of anthrax. *Stat Med.* 24(4):531–42.

83. Heppell CW, Egan JR, Hall I. A human time dose response model for Q fever. *Epidemics*. 2017 Dec;21:30–8.
84. Brauer F. Compartmental Models in Epidemiology. In: Brauer F, van den Driessche P, Wu J, editors. *Mathematical Epidemiology* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2018 Sep 24]. p. 19–79. (Lecture Notes in Mathematics). Available from: https://doi.org/10.1007/978-3-540-78911-6_2
85. Ciupe SM, Heffernan JM. In-host modeling. *Infect Dis Model*. 2017 May 1;2(2):188–202.
86. Eckhoff P. Mathematical models of within-host and transmission dynamics to determine effects of malaria interventions in a variety of transmission settings. *Am J Trop Med Hyg*. 2013 May 1;88(5):817–27.
87. Tumwiine J, Luckhaus S, Mugisha JYT, Luboobi LS. An Age-structured mathematical model for the within host dynamics of malaria and the immune system. *J Math Model Algorithms*. 2008 Mar 1;7(1):79–97.
88. Wang X, Wang J. Modeling the within-host dynamics of cholera: bacterial–viral interaction. *J Biol Dyn*. 2017 Aug 11;11(sup2):484–501.
89. Smith RJ, Li J, Mao J, Sahai B. Using within-host mathematical modelling to predict the long-term outcome of human papillomavirus vaccines. *Can Appl Math Q*. 2013 Jun;21(2):281–99.
90. Cronin P, Ryan F, Coughlan M. Undertaking a literature review: a step by step approach. *Br J Nurs*. 17(1):38–43.
91. Rozas LW, Klein WC. The value and purpose of the traditional qualitative literature review. *J Evid-Based Soc Work*. 2010 Oct;7(5):387–99.
92. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003 Mar;96(3):118.
93. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. In: *Cochrane Database of Systematic Reviews* [Internet]. John Wiley & Sons, Ltd; 2014 [cited 2016 Sep 16]. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/14651858.MR000034.pub2/abstract>
94. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
95. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998 Dec;51(12):1235–41.
96. WebPlotDigitizer - Copyright 2010-2016 Ankit Rohatgi [Internet]. [cited 2016 Jul 19]. Available from: <https://automeris.io/WebPlotDigitizer/>
97. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003 Sep 6;327(7414):557–60.

98. Sheldon MR, Fillyaw MJ, Thompson WD. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiother Res Int J Res Clin Phys Ther*. 1996;1(4):221–8.
99. Higgins J, Green S (editors). *Cochrane handbook for systematic reviews of interventions* Version 5.1.0 [updated March 2011]. Cochrane Collab [Internet]. 2011 [cited 2016 Apr 22]; Available from: www.handbook.cochrane.org
100. Kolmogorov–Smirnov Test. In: *The Concise Encyclopedia of Statistics* [Internet]. New York, NY: Springer New York; 2008 [cited 2018 Oct 24]. p. 283–7. Available from: https://doi.org/10.1007/978-0-387-32833-1_214
101. Bridges CC. Hierarchical Cluster Analysis. *Psychol Rep*. 1966 Jun 1;18(3):851–4.
102. Zhang Z, Murtagh F, Van Poucke S, Lin S, Lan P. Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. *Ann Transl Med* [Internet]. 2017 Feb [cited 2018 Jul 4];5(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5337204/>
103. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. *J Clin Epidemiol*. 1995 Jan;48(1):167–71.
104. Haidich AB. Meta-analysis in medical research. *Hippokratia*. 2010 Dec;14(Suppl 1):29–37.
105. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet Lond Engl*. 1999 Nov 27;354(9193):1896–900.
106. Hemens BJ, Iorio A. Computer-aided systematic review screening comes of age. *Ann Intern Med*. 2017 Aug 1;167(3):210.
107. Pautasso M. Ten simple rules for writing a literature review. *PLoS Comput Biol* [Internet]. 2013 Jul [cited 2016 Nov 14];9(7). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715443/>
108. Gunnarsson H, Svedhem Å. The usefulness of Diffusion-In-Gel-ELISA in clinical practice as illustrated by a *Campylobacter jejuni* outbreak. *J Immunol Methods*. 1998 Jun 1;215(1–2):135–44.
109. Morgan D, Gunneberg C, Gunnell D, Healing TD, Lamerton S, Soltanpoor N, et al. An outbreak of *Campylobacter* infection associated with the consumption of unpasteurised milk at a large festival in England. *Eur J Epidemiol*. 1994 Oct;10(5):581–5.
110. Frumkin H, Berlin J. Asbestos exposure and gastrointestinal malignancy review and meta-analysis. *Am J Ind Med*. 1988 Jan 1;14(1):79–95.
111. Silva J, Leite D, Fernandes M, Mena C, Gibbs PA, Teixeira P. *Campylobacter spp.* as a foodborne pathogen: a review. *Front Microbiol* [Internet]. 2011 [cited 2016 Oct 6];2. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2011.00200/abstract>

112. Wagenaar JA, Bergen MAP van, Blaser MJ, Tauxe RV, Newell DG, Putten JPM van. *Campylobacter fetus* infections in humans: exposure and disease. Clin Infect Dis. 2014 Feb 18;ciu085.
113. Blaser MJ, Taylor DN, Feldman RA. Epidemiology of *Campylobacter jejuni* infections. Epidemiol Rev. 1983 Jan 1;5(1):157–76.
114. Blaser MJ. Epidemiologic and clinical features of *Campylobacter jejuni* infections. J Infect Dis. 1997 Dec;176 Suppl 2:S103-105.
115. European Food Safety Authority (EFSA). The community summary report on trends and sources of zoonoses, zoonotic agents, antimicrobial resistance and foodborne outbreaks in the European Union in 2006. EFSA J. 2007 Dec 19;6:130r.
116. European Food Safety Authority (EFSA). The community summary report on trends and sources of zoonoses and zoonotic agents in the European Union in 2007. EFSA J. 2009 Jan;7(1):223r.
117. Food Standards Agency. *Campylobacter* [Internet]. [cited 2016 Oct 8]. Available from: <https://www.food.gov.uk/science/microbiology/campylobacterevidenceprogramme>
118. Evans MR, Roberts RJ, Ribeiro CD, Gardner D, Kembrey D. A milk-borne *campylobacter* outbreak following an educational farm visit. Epidemiol Infect. 1996 Dec;117(3):457.
119. Roels TH, Wickus B, Bostrom HH, Kazmierczak JJ, Nicholson MA, Kurzynski TA, et al. A foodborne outbreak of *Campylobacter jejuni* (O:33) infection associated with tuna salad: a rare strain in an unusual vehicle. Epidemiol Infect. 1998 Oct;121(2):281–7.
120. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of *Campylobacter* infection. Clin Microbiol Rev. 2015 Jul;28(3):687–720.
121. Stuart TL, Sandhu J, Stirling R, Corder J, Ellis A, Misa P, et al. Campylobacteriosis outbreak associated with ingestion of mud during a mountain bike race. Epidemiol Infect. 2010 Dec;138(12):1695–703.
122. Little CL, Gormley FJ, Rawal N, Richardson JF. A recipe for disaster: outbreaks of campylobacteriosis associated with poultry liver pâté in England and Wales. Epidemiol Infect. 2010 Dec;138(12):1691–4.
123. NHS Choices. Food poisoning - Causes - NHS Choices [Internet]. 2015 [cited 2016 Oct 12]. Available from: <http://www.nhs.uk/Conditions/Food-poisoning/Pages/Causes.aspx>
124. World Health Organization. *Campylobacter* [Internet]. WHO. [cited 2016 Jun 21]. Available from: <http://www.who.int/mediacentre/factsheets/fs255/en/>
125. Public Health Agency of Canada. *Campylobacter jejuni* - Pathogen safety data sheets [Internet]. 2012 [cited 2016 Dec 19]. Available from: <http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/campylobacter-jejuni-eng.php>
126. Horn BJ, Lake RJ. Incubation period for campylobacteriosis and its importance in the estimation of incidence related to travel. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2013;18(40).

127. R: A language and environment for statistical computing. R foundation for statistical computing. [Internet]. Vienna, Austria; 2015. Available from: [//www.R-project.org/](http://www.R-project.org/)
128. Nichols GL, Richardson JF, Sheppard SK, Lane C, Sarran C. *Campylobacter* epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open*. 2012 Jan 1;2(4):e001179.
129. Blaser MJ, Newman LS. A review of human salmonellosis: I. infective dose. *Rev Infect Dis*. 1982 Nov 1;4(6):1096–106.
130. Platts-Mills JA, Kosek M. Update on the burden of *Campylobacter* in developing countries. *Curr Opin Infect Dis*. 2014 Oct;27(5):444–50.
131. Altekruse SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*--an emerging foodborne pathogen. *Emerg Infect Dis*. 1999 Feb;5(1):28–35.
132. Bianchini V, Borella L, Benedetti V, Parisi A, Miccolupo A, Santoro E, et al. Prevalence in bulk tank milk and epidemiology of *Campylobacter jejuni* in dairy herds in Northern Italy. *Appl Environ Microbiol*. 2014 Mar;80(6):1832–7.
133. El-Sharoud WM. Prevalence and survival of *Campylobacter* in Egyptian dairy products. *Food Res Int*. 2009 Jun;42(5–6):622–6.
134. Modi S, Brahmabhatt MN, Chatur YA, Nayak JB. Prevalence of *Campylobacter* species in milk and milk products, their virulence gene profile and anti-bio gram. *Vet World*. 2015 Jan;8(1):1–8.
135. Wood RM, Egan JR, Hall IM. A dose and time response Markov model for the in-host dynamics of infection with intracellular bacteria following inhalation: with application to *Francisella tularensis*. *J R Soc Interface*. 2014 Jun 6;11(95):20140119.
136. Toth DJA, Gundlapalli AV, Schell WA, Bulmahn K, Walton TE, Woods CW, et al. Quantitative models of the dose-response and time course of inhalational anthrax in humans. *PLoS Pathog* [Internet]. 2013 Aug [cited 2016 Oct 31];9(8). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3744436/>
137. Aho M, Kurki M, Rautelin H, Kosunen TU. Waterborne outbreak of *Campylobacter* enteritis after outdoors infantry drill in Utti, Finland. *Epidemiol Infect*. 1989 Aug;103(1):133.
138. World Health Organization. WHO | *E. coli* [Internet]. WHO. [cited 2017 Nov 1]. Available from: <http://www.who.int/mediacentre/factsheets/fs125/en/>
139. Thorpe CM. Shiga Toxin—Producing *Escherichia coli* Infection. *Clin Infect Dis*. 2004 May 1;38(9):1298–303.
140. Centre for Disease Prevention and Control. General Information| *E.coli* | CDC [Internet]. [cited 2017 Nov 1]. Available from: <https://www.cdc.gov/ecoli/general/index.html>
141. Byrne L, Jenkins C, Launder N, Elson R, Adak GK. The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol Infect*. 2015 Dec;143(16):3475–87.

142. Karch H, Bielaszewska M, Bitzan M, Schmidt H. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Diagn Microbiol Infect Dis*. 1999 Jul;34(3):229–43.
143. Smith JL, Fratamico PM, Gunther NW. Chapter Three - Shiga Toxin-producing *Escherichia coli*. In: Sariaslani S, Gadd GM, editors. *Advances in Applied Microbiology* [Internet]. Academic Press; 2014 [cited 2017 Nov 29]. p. 145–97. Available from: <http://www.sciencedirect.com/science/article/pii/B9780128002629000032>
144. Byrne L, Vanstone GL, Perry NT, Launders N, Adak GK, Godbole G, et al. Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013. *J Med Microbiol*. 2014 Sep;63(Pt 9):1181–8.
145. Hadler JL, Clogher P, Hurd S, Phan Q, Mandour M, Bemis K, et al. Ten-year trends and risk factors for non-O157 Shiga toxin-producing *Escherichia coli* Found Through Shiga Toxin Testing, Connecticut, 2000–2009. *Clin Infect Dis*. 2011 Aug 1;53(3):269–76.
146. Parry SM, Salmon RL. Sporadic STEC O157 infection: secondary household transmission in Wales. *Emerg Infect Dis*. 1998 Dec;4(4):657–61.
147. Adams NL, Byrne L, Smith GA, Elson R, Harris JP, Salmon R, et al. Shiga toxin-producing *Escherichia coli* O157, England and Wales, 1983–2012. *Emerg Infect Dis*. 2016 Apr;22(4):590–7.
148. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, et al. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 Outbreak in Germany. *N Engl J Med*. 2011 Nov 10;365(19):1771–80.
149. Ihekweazu C, Carroll K, Adak B, Smith G, Pritchard GC, Gillespie IA, et al. Large outbreak of verocytotoxin-producing *Escherichia coli* O157 infection in visitors to a petting farm in South East England, 2009. *Epidemiol Infect*. 2012 Aug;140(8):1400–13.
150. Devaux I, Varela-Santos C, Payne-Hallström L, Takkinen J, Bogaardt C, Coulombier D. Investigation of travel-related cases in a multinational outbreak: example of the Shiga-toxin producing *E. coli* outbreak in Germany, May–June 2011. *Epidemiol Infect*. 2015 Dec;143(16):3468–74.
151. Awofisayo-Okuyelu A, Hall I, Adak G, Hawker JI, Abbott S, McCarthy N. A systematic review and meta-analysis on the incubation period of Campylobacteriosis. *Epidemiol Infect*. 2017 Aug;145(11):2241–53.
152. Scharlach M, Diercke M, Dreesman J, Jahn N, Kriek M, Beyrer K, et al. Epidemiological analysis of a cluster within the outbreak of Shiga toxin-producing *Escherichia coli* serotype O104:H4 in Northern Germany, 2011. *Int J Hyg Environ Health*. 2013 Jun;216(3):341–5.
153. Uphoff H, Hedrich B, Strotmann I, Arvand M, Bettge-Weller G, Hauri AM. A prolonged investigation of an STEC-O104 cluster in Hesse, Germany, 2011 and implications for outbreak management. *J Public Health*. 2014 Feb 1;22(1):41–8.
154. Diercke M, Kirchner M, Claussen K, Mayr E, Strotmann I, Frangenberg J, et al. Transmission of shiga toxin-producing *Escherichia coli* O104:H4 at a family party

- possibly due to contamination by a food handler, Germany 2011. *Epidemiol Amp Infect.* 2014 Jan;142(1):99–106.
155. Werber D, King LA, Müller L, Follin P, Buchholz U, Bernard H, et al. Associations of age and sex with the clinical outcome and incubation period of Shiga toxin-producing *Escherichia coli* O104:H4 infections, 2011. *Am J Epidemiol.* 2013 Sep 15;178(6):984–92.
 156. Glynn JR, Bradley DJ. The relationship between infecting dose and severity of disease in reported outbreaks of Salmonella infections. *Epidemiol Infect.* 1992 Dec;109(3):371–88.
 157. Hornick RB, Music SI, Wenzel R, Cash R, Libonati JP, Snyder MJ, et al. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bull N Y Acad Med.* 1971 Oct;47(10):1181–91.
 158. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Determination of the 50% human infectious dose for norwalk virus. *J Infect Dis.* 2014 Apr 1;209(7):1016–22.
 159. Lee RM, Lessler J, Lee RA, Rudolph KE, Reich NG, Perl TM, et al. Incubation periods of viral gastroenteritis: a systematic review. *BMC Infect Dis.* 2013;13:446.
 160. Virlogeux V, Fang VJ, Wu JT, Ho L-M, Malik Peiris JS, Leung GM, et al. Incubation period duration and severity of clinical disease following severe acute respiratory syndrome coronavirus infection. *Epidemiol Camb Mass.* 2015 Sep;26(5):666–9.
 161. Characteristics of Infectious Disease | Microbiology [Internet]. [cited 2017 Dec 30]. Available from: <https://courses.lumenlearning.com/microbiology/chapter/characteristics-of-infectious-disease/>
 162. King LA, Nogareda F, Weill F-X, Mariani-Kurkdjian P, Loukiadis E, Gault G, et al. Outbreak of Shiga toxin-producing *Escherichia coli* O104:H4 associated with organic fenugreek sprouts, France, June 2011. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 2012 Jun;54(11):1588–94.
 163. Vally H, Hall G, Dyda A, Raupach J, Knope K, Combs B, et al. Epidemiology of Shiga toxin producing *Escherichia coli* in Australia, 2000-2010. *BMC Public Health.* 2012 Jan 21;12:63.
 164. Cody SH, Glynn MK, Farrar JA, Cairns KL, Griffin PM, Kobayashi J, et al. An outbreak of *Escherichia coli* O157:H7 infection from unpasteurized commercial apple juice. *Ann Intern Med.* 1999 Feb 2;130(3):202–9.
 165. Ferguson DD, Scheftel J, Cronquist A, Smith K, Woo-Ming A, Anderson E, et al. Temporally distinct *Escherichia coli* O157 outbreaks associated with alfalfa sprouts linked to a common seed source--Colorado and Minnesota, 2003. *Epidemiol Infect.* 2005 Jun;133(3):439–47.
 166. Ackman D, Marks S, Mack P, Caldwell M, Root T, Birkhead G. Swimming-associated haemorrhagic colitis due to *Escherichia coli* O157:H7 infection: evidence of prolonged contamination of a fresh water lake. *Epidemiol Infect.* 1997 Aug;119(1):1–8.

167. Aldabe B, Delmas Y, Gault G, Vendrely B, Llanas B, Charron M, et al. Household transmission of haemolytic uraemic syndrome associated with *Escherichia coli* O104:H4, south-western France, June 2011. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2011 Aug 4;16(31).
168. Razzaq S. Hemolytic Uremic Syndrome: An Emerging Health Risk. Am Fam Physician. 2006 Sep 15;74(6):991–6.
169. Launders N, Byrne L, Jenkins C, Harker K, Charlett A, Adak GK. Disease severity of Shiga toxin-producing *E. coli* O157 and factors influencing the development of typical haemolytic uraemic syndrome: a retrospective cohort study, 2009–2012. BMJ Open. 2016 Jan 1;6(1):e009933.
170. Public Health England. Shiga toxin-producing *Escherichia coli*: questionnaire [Internet]. GOV.UK. [cited 2018 Jul 4]. Available from: <https://www.gov.uk/government/publications/vero-cytotoxin-producing-escherichia-coli-questionnaire>
171. Cleary TG. The role of Shiga-toxin-producing *Escherichia coli* in hemorrhagic colitis and hemolytic uremic syndrome. Semin Pediatr Infect Dis. 2004 Oct;15(4):260–5.
172. Cohen MB, Giannella RA. Hemorrhagic colitis associated with *Escherichia coli* O157:H7. Adv Intern Med. 1992;37:173–95.
173. Mintz ED, Cartter ML, Hadler JL, Wassell JT, Zingales JA, Tauxe RV. Dose-Response Effects in an Outbreak of *Salmonella* enteritidis. Epidemiol Infect. 1994;112(1):13–23.
174. Paton JC, Paton AW. Pathogenesis and Diagnosis of Shiga toxin-producing *Escherichia coli* Infections. Clin Microbiol Rev. 1998 Jul;11(3):450–79.
175. Zumbrun SD, Melton-Celsa AR, Smith MA, Gilbreath JJ, Merrell DS, O'Brien AD. Dietary choice affects Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 colonization and disease. Proc Natl Acad Sci U S A. 2013 Jun 4;110(23):E2126–33.
176. Dallman TJ, Byrne L, Ashton PM, Cowley LA, Perry NT, Adak G, et al. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. Clin Infect Dis Off Publ Infect Dis Soc Am. 2015 Aug 1;61(3):305–12.
177. Milne LM, Plom A, Strudley I, Pritchard GC, Crooks R, Hall M, et al. *Escherichia coli* O157 incident associated with a farm open to members of the public. Commun Dis Public Health. 1999 Jan;2(1):22–6.
178. Bell BP, Goldoft M, Griffin PM, Davis MA, Gordon DC, Tarr PI, et al. A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. JAMA. 1994 Nov 2;272(17):1349–53.
179. Michino H, Araki K, Minami S, Takaya S, Sakai N, Miyazaki M, et al. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am J Epidemiol. 1999 Oct 15;150(8):787–96.
180. Pang T. Typhoid fever research in developing countries. J Infect Dev Ctries. 2008 Dec 1;2(6):411.

181. Tsang R, Wong A, Li K. Identification and serotyping of *Salmonella* typhi. Lab Med. 1989 Nov;767–71.
182. Dougan G, Baker S. *Salmonella* enterica serovar Typhi and the pathogenesis of typhoid fever. Annu Rev Microbiol. 2014;68:317–36.
183. Greenwell J, McCool J, Kool J, Salusalu M. Typhoid fever: hurdles to adequate hand washing for disease prevention among the population of a peri-urban informal settlement in Fiji. West Pac Surveill Response J WPSAR. 2013 Jan;4(1):41–5.
184. Jin C, Gibani MM, Moore M, Juel HB, Jones E, Meiring J, et al. Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of *Salmonella* Typhi: a randomised controlled, phase 2b trial. The Lancet [Internet]. 2017 Sep 28 [cited 2017 Oct 10];0(0). Available from: [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)32149-9/abstract](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32149-9/abstract)
185. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid Fever. N Engl J Med. 2002 Nov 28;347(22):1770–82.
186. Mogasale V, Maskery B, Ochiai RL, Lee JS, Mogasale VV, Ramani E, et al. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. Lancet Glob Health. 2014 Oct;2(10):e570-580.
187. Sawyer WA. Ninety-three persons infected by a typhoid carrier at a public dinner. J Am Med Assoc. 1914 Oct 31;LXIII(18):1537–42.
188. Centre for Disease Prevention and Control. Chapter 3: Infectious diseases related to travel - Typhoid & Paratyphoid fever. In: CDC Yellow Book 2018 [Internet]. Oxford University Press; 2017 [cited 2017 Oct 10]. Available from: <https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/typhoid-paratyphoid-fever>
189. Cumming JG. An epidemic resulting from the contamination of ice cream by a typhoid carrier. J Am Med Assoc. 1917 Apr 21;LXVIII(16):1163–5.
190. Freedman J, Lighton L, Jones J. Defining travel-associated cases of enteric fever. J Infect Public Health. 2014 Sep 1;7(5):377–85.
191. Raffatellu M, Wilson RP, Winter SE, Bäumler AJ. Clinical pathogenesis of typhoid fever. J Infect Dev Ctries. 2008 Aug 30;2(4):260–6.
192. Hornick RB, Greisman SE, Woodward TE, DuPont HL, Dawkins AT, Snyder MJ. Typhoid Fever: Pathogenesis and Immunologic Control. N Engl J Med. 1970 Sep 24;283(13):686–91.
193. Naylor GR. Incubation period and other features of food-borne and water-borne outbreaks of typhoid fever in relation to pathogenesis and genetics of resistance. Lancet Lond Engl. 1983 Apr 16;1(8329):864–6.
194. Choices NHS. Typhoid fever - Vaccination - NHS Choices [Internet]. 2017 [cited 2017 Oct 10]. Available from: <http://www.nhs.uk/Conditions/Typhoid-fever/Pages/Prevention.aspx>

195. Vaccine Information Statement | Typhoid | VIS | CDC [Internet]. [cited 2017 Oct 10]. Available from: <https://www.cdc.gov/vaccines/hcp/vis/vis-statements/typhoid.html>
196. Caraway CT, Bruce JM. Typhoid fever epidemic following a wedding reception. Public Health Rep. 1961 May;76(5):427–30.
197. Michel R, Garnotel E, Spiegel A, Morillon M, Saliou P, Boutin J-P. Outbreak of typhoid fever in vaccinated members of the French Armed Forces in the Ivory Coast. Eur J Epidemiol. 2005 Jul 1;20(7):635–42.
198. Bajracharya D, Khan MI, Iii AP, Shrestha P, Joshi N, Upreti SR, et al. 25 Years after Vi Typhoid Vaccine Efficacy Study, Typhoid Affects Significant Number of Population in Nepal. PLOS ONE. 2014 Jan 6;9(1):e77974.
199. Bayram Y, Gdcođlu H, Otl B, Aypak C, Grsoy NC, Uluç H, et al. Epidemiological characteristics and molecular typing of *Salmonella* enterica serovar Typhi during a waterborne outbreak in Eastern Anatolia. Ann Trop Med Parasitol. 2011 Jul;105(5):359–65.
200. Centers for Disease Control and Prevention (CDC). Notes from the field: *Salmonella* Typhi infections associated with contaminated water--Zimbabwe, October 2011-May 2012. MMWR Morb Mortal Wkly Rep. 2012 Jun 15;61(23):435.
201. Egoz N, Shihab S, Leitner L, Lucian M. An outbreak of typhoid fever due to contamination of the municipal water supply in northern Israel. Isr J Med Sci. 1988 Nov;24(11):640–3.
202. Roy JS, Saikia L, Medhi M, Tassa D. Epidemiological investigation of an outbreak of typhoid fever in Jorhat town of Assam, India. Indian J Med Res. 2016 Oct;144(4):592–6.
203. Loharikar A, Newton A, Rowley P, Wheeler C, Bruno T, Barillas H, et al. Typhoid fever outbreak associated with frozen mamey pulp imported from Guatemala to the western United States, 2010. Clin Infect Dis Off Publ Infect Dis Soc Am. 2012 Jul;55(1):61–6.
204. Taylor A, Santiago A, Gonzalez-Cortes A, Gangarosa EJ. Outbreak of typhoid fever in Trinidad in 1971 traced to a commercial ice cream product. Am J Epidemiol. 1974 Aug;100(2):150–7.
205. al-Zubaidy AA, el Bushra HE, Mawlawi MY. An outbreak of typhoid fever among children who attended a potluck dinner at Al-Mudhnab, Saudi Arabia. East Afr Med J. 1995 Jun;72(6):373–5.
206. Ash I, Mckendrick GD, Robertson MH, Hughes HL. Outbreak of typhoid fever connected with corned beef. Br Med J. 1964 Jun 6;1(5396):1474–8.
207. Glynn JR, Hornick RB, Levine MM, Bradley DJ. Infecting dose and severity of typhoid: analysis of volunteer data and examination of the influence of the definition of illness used. Epidemiol Infect. 1995 Aug;115(1):23–30.
208. Pradier C, Keita-Perse O, Bernard E, Gisbert C, Vezolles MJ, Armengaud A, et al. Outbreak of typhoid fever on the French Riviera. Eur J Clin Microbiol Infect Dis Off Publ Eur Soc Clin Microbiol. 2000 Jun;19(6):464–7.

209. Anita S, Amir KM, Fadzilah K, Ahamad J, Noorhaida U, Marina K, et al. Risk factors for typhoid outbreak in Sungai Congkak Recreational Park, Selangor 2009. *Med J Malaysia*. 2012 Feb;67(1):12–6.
210. Coté TR, Convery H, Robinson D, Ries A, Barrett T, Frank L, et al. Typhoid fever in the park: epidemiology of an outbreak at a cultural interface. *J Community Health*. 1995 Dec;20(6):451–8.
211. Judd MC, Mintz ED. Chapter 3. Infectious diseases related to travel: Typhoid and paratyphoid. In: *Yellow Book* [Internet]. [cited 2018 Jul 28]. Available from: <https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/typhoid-paratyphoid-fever>
212. Hunter J, Watkins L. Chapter 3. Infectious diseases related to travel: Salmonellosis (Nontyphoidal). In: *Yellow Book* [Internet]. [cited 2018 Jul 28]. Available from: <https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/salmonellosis-nontyphoidal>
213. Glynn JR, Palmer SR. Incubation period, severity of disease, and infecting dose: evidence from a *Salmonella* outbreak. *Am J Epidemiol*. 1992 Dec 1;136(11):1369–77.
214. Bhutta ZA. Typhoid Fever: Current Concepts. *Infect Dis Clin Pract*. 2006 Sep;14(5):266.
215. Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive *Salmonella* infections. *Clin Microbiol Rev*. 2015 Oct 1;28(4):901–37.
216. Cordova-Fraga T, Sosa M, Wiechers C, la Roca-Chiapas JMD, Moreles AM, Bernal-Alvarado J, et al. Effects of anatomical position on esophageal transit time: A biomagnetic diagnostic technique. *World J Gastroenterol WJG*. 2008 Oct 7;14(37):5707–11.
217. Maurer AH. Gastrointestinal Motility, Part 1: Esophageal transit and Gastric emptying. *J Nucl Med*. 2015 Aug 1;56(8):1229–38.
218. Kondo H, Watanabe T, Yokohama S, Watanabe J. Effect of food on gastrointestinal transit of liquids in cynomolgus monkeys. *Biopharm Drug Dispos*. 2003 May;24(4):141–51.
219. Vasavid P, Chaiwatanarat T, Pusuwan P, Sritara C, Roysri K, Namwongprom S, et al. Normal solid gastric emptying values measured by scintigraphy using asian-style meal: a multicenter study in healthy volunteers. *J Neurogastroenterol Motil*. 2014 Jul;20(3):371–8.
220. Bennink R, Peeters M, Van den Maegdenbergh V, Geypens B, Rutgeerts P, De Roo M, et al. Evaluation of small-bowel transit for solid and liquid test meal in healthy men and women. *Eur J Nucl Med*. 1999 Dec;26(12):1560–6.
221. Hellmig S, Von Schöning F, Gadow C, Katsoulis S, Hedderich J, Fölsch UR, et al. Gastric emptying time of fluids and solids in healthy subjects determined by ¹³C breath tests: influence of age, sex and body mass index. *J Gastroenterol Hepatol*. 2006 Dec;21(12):1832–8.

222. Szarka LA, Camilleri M. Methods for the assessment of small bowel and colonic transit. *Semin Nucl Med.* 2012 Mar;42(2):113–23.
223. Degen LP, Phillips SF. Variability of gastrointestinal transit in healthy women and men. *Gut.* 1996 Aug;39(2):299–305.
224. Argenyi EE, Soffer EE, Madsen MT, Berbaum KS, Walkner WO. Scintigraphic evaluation of small bowel transit in healthy subjects: inter- and intrasubject variability. *Am J Gastroenterol.* 1995 Jun;90(6):938–42.
225. Grybäck P, Jacobsson H, Blomquist L, Schnell PO, Hellström PM. Scintigraphy of the small intestine: a simplified standard for study of transit with reference to normal values. *Eur J Nucl Med Mol Imaging.* 2002 Jan;29(1):39–45.
226. Yu LX, Crison JR, Amidon GL. Compartmental transit and dispersion model analysis of small intestinal transit flow in humans. *Int J Pharm.* 1996 Aug 16;140(1):111–8.
227. Kaiser P, Slack E, Grant AJ, Hardt W-D, Regoes RR. Lymph Node colonization dynamics after oral *Salmonella* typhimurium infection in mice. *PLOS Pathog.* 2013 Sep 19;9(9):e1003532.
228. Hallstrom K, McCormick BA. *Salmonella* interaction with and passage through the intestinal mucosa: through the lens of the organism. *Front Microbiol* [Internet]. 2011 Apr 29 [cited 2017 Apr 3];2. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3128981/>
229. Knodler LA, Vallance BA, Celli J, Winfree S, Hansen B, Montero M, et al. Dissemination of invasive *Salmonella* via bacterial-induced extrusion of mucosal epithelia. *Proc Natl Acad Sci U S A.* 2010 Oct 12;107(41):17733–8.
230. Zhang S, Kingsley RA, Santos RL, Andrews-Polymenis H, Raffatellu M, Figueiredo J, et al. Molecular pathogenesis of *Salmonella* enterica serotype typhimurium-induced diarrhea. *Infect Immun.* 2003 Jan;71(1):1–12.
231. Mastroeni P, Grant AJ. Spread of *Salmonella* enterica in the body during systemic infection: unravelling host and pathogen determinants. *Expert Rev Mol Med.* 2011 Apr 11;13:e12.
232. Gog JR, Murcia A, Osterman N, Restif O, McKinley TJ, Sheppard M, et al. Dynamics of *Salmonella* infection of macrophages at the single cell level. *J R Soc Interface.* 2012 Oct 7;9(75):2696–707.
233. Hueffer K, Galán JE. *Salmonella*-induced macrophage death: multiple mechanisms, different outcomes. *Cell Microbiol.* 2004 Nov;6(11):1019–25.
234. Fenlon LA, Slauch JM. Phagocyte roulette in *Salmonella* killing. *Cell Host Microbe.* 2014 Jan 15;15(1):7–8.
235. Eng S-K, Pusparajah P, Mutalib N-SA, Ser H-L, Chan K-G, Lee L-H. *Salmonella*: A review on pathogenesis, epidemiology and antibiotic resistance. *Front Life Sci.* 2015 Jul 3;8(3):284–93.
236. Forest CG, Ferraro E, Sabbagh SC, Daigle F. Intracellular survival of *Salmonella* enterica serovar Typhi in human macrophages is independent of *Salmonella* pathogenicity island (SPI)-2. *Microbiology.* 2010;156(12):3689–98.

237. Lowrie DB, Aber VR, Carrol ME. Division and death rates of *Salmonella* typhimurium inside macrophages: use of penicillin as a probe. J Gen Microbiol. 1979 Feb;110(2):409–19.
238. Monack DM, Raupach B, Hromockyj AE, Falkow S. *Salmonella* typhimurium invasion induces apoptosis in infected macrophages. Proc Natl Acad Sci U S A. 1996 Sep 3;93(18):9833–8.
239. Schwan WR, Huang XZ, Hu L, Kopecko DJ. Differential bacterial survival, replication, and apoptosis-inducing ability of *Salmonella* serovars within human and murine macrophages. Infect Immun. 2000 Mar;68(3):1005–13.
240. Monack DM, Navarre WW, Falkow S. *Salmonella*-induced macrophage death: the role of caspase-1 in death and inflammation. Microbes Infect. 2001 Dec;3(14–15):1201–12.
241. Tam MA, Rydström A, Sundquist M, Wick MJ. Early cellular responses to *Salmonella* infection: dendritic cells, monocytes, and more. Immunol Rev. 2008 Oct 1;225(1):140–62.
242. Freitas Jr. R. Nanomedicine, Vol. I: Basic Capabilities [Internet]. Vol. I. Georgetown, TX: Landes Bioscience; 1999 [cited 2017 Mar 10]. Available from: <http://www.nanomedicine.com/NMI/8.2.1.3.htm>
243. Alexander JS, Ganta VC, Jordan PA, Witte MH. Gastrointestinal lymphatics in health and disease. Pathophysiol Off J Int Soc Pathophysiol ISP. 2010 Sep;17(4):315–35.
244. Son G, Kremer M, Hines IN. Contribution of gut bacteria to liver pathobiology [internet]. gastroenterology research and practice. 2010 [cited 2017 Jul 14]. Available from: <https://www.hindawi.com/journals/grp/2010/453563/>
245. Brown HS, Halliwell M, Qamar M, Read AE, Evans JM, Wells PN. Measurement of normal portal venous blood flow by Doppler ultrasound. Gut. 1989 Apr;30(4):503–9.
246. Perisić MD, Culafić DM, Kerkez M. Specificity of splenic blood flow in liver cirrhosis. Romanian J Intern Med Rev Roum Med Interne. 2005;43(1–2):141–51.
247. Sato S, Ohnishi K, Sugita S, Okuda K. Splenic artery and superior mesenteric artery blood flow: nonsurgical Doppler US measurement in healthy subjects and patients with chronic liver disease. Radiology. 1987 Aug;164(2):347–52.
248. Watson KG, Holden DW. Dynamics of growth and dissemination of *Salmonella* in vivo. Cell Microbiol. 2010 Oct 1;12(10):1389–97.
249. Harris JR, Bergmire-Sweat D, Schlegel JH, Winpisinger KA, Klos RF, Perry C, et al. Multistate outbreak of *Salmonella* infections associated with small turtle exposure, 2007–2008. Pediatrics. 2009 Nov;124(5):1388–94.
250. Hoffman TA, Ruiz CJ, Counts GW, Sachs JM, Nitzkin JL. Waterborne typhoid fever in Dade County, Florida. Clinical and therapeutic evaluation of 105 bacteremic patients. Am J Med. 1975 Oct;59(4):481–7.
251. Moore WB. Typhoid Fever, with particular reference to the Crowthorne epidemic, 1949*. J R Sanit Inst. 1950 Mar 1;70(2):93–101.

252. Ramsey GH, Benning CH, Orr PF. An epidemic of typhoid fever following a church dinner. *Am J Public Health N Y N* 1912. 1926 Oct;16(10):1011–6.
253. Farooqui A, Khan A, Kazmi SU. Investigation of a community outbreak of typhoid fever associated with drinking water. *BMC Public Health*. 2009 Dec 20;9:476.
254. Yu LX, Amidon GL. Characterization of small intestinal transit time distribution in humans. *Int J Pharm*. 1998 Sep 15;171(2):157–63.
255. Dissel JT van, Leijh PC, Furth R van. Differences in initial rate of intracellular killing of *Salmonella typhimurium* by resident peritoneal macrophages from various mouse strains. *J Immunol*. 1985 May 1;134(5):3404–10.
256. Mitsuhashi S, Sato I, Tanaka T. Experimental salmonellosis. Intracellular growth of *Salmonella enteritidis* ingested in mononuclear phagocytes of mice, and cellular basis of immunity. *J Bacteriol*. 1961 Jun;81:863–8.
257. Hoare A, Regan DG, Wilson DP. Sampling and sensitivity analyses tools (SaSAT) for computational modelling. *Theor Biol Med Model*. 2008 Feb 27;5:4.
258. Shields MD, Zhang J. The generalization of Latin hypercube sampling. *Reliab Eng Syst Saf*. 2016 Apr 1;148:96–108.
259. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J Theor Biol*. 2008 Sep 7;254(1):178–96.
260. Tennant SM, Hartland EL, Phumoonna T, Lyras D, Rood JI, Robins-Browne RM, et al. Influence of Gastric Acid on Susceptibility to Infection with Ingested Bacterial Pathogens. *Infect Immun*. 2008 Feb 1;76(2):639–45.
261. Giannella RA, Broitman SA, Zamcheck N. Gastric acid barrier to ingested microorganisms in man: studies in vivo and in vitro. *Gut*. 1972 Apr;13(4):251–6.
262. Bruschi J. Typhoid Fever follow-up: further outpatient care, further inpatient care, deterrence/prevention. 2017 Sep 19; Available from: <https://emedicine.medscape.com/article/231135-followup#e1>
263. Gibani MM, Jin C, Darton TC, Pollard AJ. Control of invasive *Salmonella* disease in Africa: Is There a Role for Human Challenge Models? *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2015 Nov 1;61 Suppl 4:S266–271.
264. Jin C, Gibani MM, Moore M, Juel HB, Jones E, Meiring J, et al. Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of *Salmonella Typhi*: a randomised controlled, phase 2b trial. *The Lancet*. 2017 Dec 2;390(10111):2472–80.
265. Wain J, Diep TS, Ho VA, Walsh AM, Hoa NTT, Parry CM, et al. Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *J Clin Microbiol*. 1998 Jun;36(6):1683–7.
266. Tennant S, Toema D, Qamar F, Iqbal N, Boyd M, Marshall J, et al. Detection of typhoidal and paratyphoidal *Salmonella* in blood by real-time polymerase chain reaction. *Clin Infect Dis*. 2015 Nov 1;61(4):S241–50.

267. Torin DE, Betts SL, McClenahan JB, Phillips GM, Shikamura MT. A typhoid fever outbreak on a university campus. *Arch Intern Med.* 1969 Nov;124(5):606–10.
268. Sartwell PE. The distribution of incubation periods of infectious disease. *Am J Hyg.* 1950 May;51(3):310–8.
269. Ziegel E. Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory. *Technometrics.* 1989 Feb 1;31(1):128–128.
270. Delignette-Muller M, Dutang C. fitdistrplus: An R package for fitting distributions. *J Stat Softw.* 64(4):1–34.
271. Bowman KO, Shenton LR. The beta distribution, moment method, Karl Pearson and R.A. Fisher. Vol. 23. 2007. 133 p.
272. Gopalakrishnan S, Ganeshkumar P. Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *J Fam Med Prim Care.* 2013;2(1):9–14.
273. Tugwell P, Knottnerus JA. Advantages of individual patient data analysis in systematic reviews. *J Clin Epidemiol.* 2010 Mar 1;63(3):233–4.
274. UK Caldicott Guardian Council [Internet] [cited 2018 Sep 17]. Available from: <https://www.gov.uk/government/groups/uk-caldicott-guardian-council>
275. WHO | Water-related diseases [Internet]. WHO. [cited 2016 Jun 21]. Available from: http://www.who.int/water_sanitation_health/diseases/campylobacteriosis/en/
276. Moffatt CRM, Greig A, Valcanis M, Gao W, Seemann T, Howden BP, et al. A large outbreak of *Campylobacter jejuni* infection in a university college caused by chicken liver pâté, Australia, 2013. *Epidemiol Infect.* 2016 Jun;FirstView:1–8.
277. Infectious disease *Campylobacter* general information | Foodborne illnesses | CDC [Internet]. [cited 2016 Jun 21]. Available from: <http://www.cdc.gov/foodsafety/diseases/campylobacter/index.html>
278. Food Standards Agency. *Campylobacter* [Internet]. [cited 2016 Jun 21]. Available from: <https://www.food.gov.uk/science/microbiology/campylobacterevidenceprogramme>
279. Unicomb LE, Fullerton KE, Kirk MD, Stafford RJ. Outbreaks of *Campylobacteriosis* in Australia, 2001 to 2006. *Foodborne Pathog Dis.* 2009 Nov 6;6(10):1241–50.
280. Inns T, Foster K, Gorton R. Cohort study of a campylobacteriosis outbreak associated with chicken liver parfait, United Kingdom, June 2010. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull.* 2010;15(44).
281. Jones AK, Rigby D, Burton M, Millman C, Williams NJ, Jones TR, et al. Restaurant cooking trends and increased risk for *Campylobacter* infection. *Emerg Infect Dis.* 2016 Jul;22(7):1208–15.
282. Hope KG, Merritt TD, Durrheim DN. Short incubation periods in *Campylobacter* outbreaks associated with poultry liver dishes. *Commun Dis Intell Q Rep.* 2014 Mar;38(1):E20-23.
283. Winkvist AG, Roome A, Mshar R, Fiorentino T, Mshar P, Hadler J. Outbreak of campylobacteriosis at a senior center. *J Am Geriatr Soc.* 2001 Mar 1;49(3):304–7.

284. Riordan T, Humphrey TJ, Fowles A. A point source outbreak of *Campylobacter* infection related to bird-pecked milk. *Epidemiol Infect.* 1993 Apr;110(2):261.
285. Korlath JA, Osterholm MT, Judy LA, Forfang JC, Robinson RA. A point-source outbreak of campylobacteriosis associated with consumption of raw milk. *J Infect Dis.* 1985 Sep 1;152(3):592–6.
286. Moffatt CRM, Cameron S, Mickan L, Givney RC. *Campylobacter jejuni* gastroenteritis at an Australian boarding school: consistency between epidemiology, *flaA* typing, and multilocus sequence typing. *Foodborne Pathog Dis.* 2010 Jul 9;7(11):1285–90.

APPENDICES

APPENDIX 1: SYSTEMATIC REVIEW OF THE INCUBATION PERIOD OF CAMPYLOBACTERIOSIS – PROTOCOL

Background

Campylobacteriosis is an intestinal infection caused by the Gram-negative bacterium, *Campylobacter* spp. It is one of the most common causes of human bacterial infection and the most common cause of gastrointestinal infection worldwide (124) . Although symptoms experienced are usually mild, the infection can be fatal among children, elderly and people who are immunocompromised.

Following infection with the bacteria, the most commonly reported symptoms are diarrhoea, abdominal cramps, nausea and vomiting, with or without fever. Symptoms usually last for about three to six days (113,114). Furthermore, the possibility of long-term sequelae resulting in severe neurological disorders following infection makes campylobacteriosis a very important infection.

According to the World Health Organization, approximately 5-15% of all diarrhoeal illness occurring worldwide is caused by *Campylobacter* (275). In developed countries like Australia and the USA, reported incidence rates are 101 cases and 14 cases per 100000 people respectively (276,277). Estimated costs to the United Kingdom economy is approximately £900 million annually (278). In developing countries, *Campylobacter* is said to result in more diarrhoeal illnesses than foodborne *Salmonella* (124).

Identified outbreaks of campylobacteriosis are uncommon(279) with the majority of the cases occurring sporadically, however, outbreaks are increasingly being reported and majority have been linked to the consumption of poultry liver pâté (280–282). Outbreak investigation is central to public health practice in order to interrupt the spread of disease by eliminating the source, as well as contributing to the body of evidence-based medicine. An important factor in the investigation of an outbreak is knowledge of the disease process of

the pathogen including the incubation period, which is the time period from infection to onset of symptoms.

In the investigation of outbreaks due to *Campylobacter*, the incubation period used by studies to develop a case definition and select a time frame for investigation varies. Some studies have used two to five days (283), while others used three to five days (284). Additionally, various statistical summaries have been reported including means of 68 hours (285) and four days (286) and medians of 24 hours (282) and ten days (60). This inconsistency means that the actual incubation time of *Campylobacter* is uncertain and this could affect implementation of public health interventions.

This study aims to systematically review scientific literature in order to identify and collate reported incubation times, synthesise the data and identify factors that may influence the incubation period.

Research question and modified PICO elements

Objective 1: What is the distribution of incubation period of campylobacteriosis?

Objective 2: What factors influence the incubation period distribution of campylobacteriosis?

Table 1. Modified PICO elements for campylobacteriosis review protocol

Population studied/Participants	Laboratory confirmed cases of <i>Campylobacter</i> spp. Probable cases of campylobacter based on clinical symptoms and case definitions in the context of outbreaks
Infectious agent	<i>Campylobacter</i> spp. <i>Campylobacter jejuni</i> <i>Campylobacter coli</i> <i>Campylobacter fetus</i>

Route of Infection	Foodborne and non-foodborne
Outcome	Time from exposure to onset of clinical illness as described or defined by the authors (diarrhoea, vomiting, nausea, abdominal cramps etc.)

Study design

Data sources and search strategy

The strategy to be used will involve systematically searching electronic databases for peer reviewed publications of observational studies or experimental studies and the reference lists of identified review papers in order to identify relevant studies.

Electronic databases

Three main databases will be searched for peer reviewed publications and conference abstracts:

- i) PubMed – one of the largest and most accessible databases of biomedical literature, life science journals and online books
- ii) Google Scholar – a search engine for scholarly literature including articles, books, abstracts etc.
- iii) ISI Web of Knowledge – a database of over one billion searchable and cited references.

Reference list

The reference list of review papers will be screened to identify other necessary studies where incubation period of *Campylobacter* spp. may have been reported.

Search terms

The following compound search term will be used: *Campylobacter* AND Humans AND Outbreaks AND/OR Experimental.

Search strings and MeSH terms

Campylobacter:

"campylobacter"[MeSH Terms] OR "campylobacter"[All Fields]

Humans:

"humans"[MeSH Terms] OR "humans"[All Fields]

Outbreaks:

"epidemiology"[Subheading] OR "epidemiology"[All Fields] OR "outbreaks"[All Fields] OR "disease outbreaks"[MeSH Terms] OR ("disease"[All Fields] AND "outbreaks"[All Fields]) OR "disease outbreaks"[All Fields]

Experimental:

experimental [All Fields]

All studies reporting infection of campylobacteriosis, either observational or experimental, will be included for screening. There will be no restriction on the reported species. In order to ensure that most studies have been included, the selection and assessment of studies will be done concurrently in the following phases:

- 5) Screening of titles and abstracts for articles with human campylobacteriosis
- 6) Screening of full text for 'incubation period'
- 7) Further review of full text to assess the reporting of incubation period.

Full texts will be downloaded and stored in either HTML formats or PDF formats. Hard copies will be requested from University of Oxford library where an electronic copy is unavailable. Where neither version (electronic or hard copies) are available, the reference will be excluded. All references will be stored and managed using the bibliographic software Zotero.

Selection criteria

The initial screening of titles, abstracts and full texts (phases 1 and 2) will be done by the main reviewer. Articles will be included or excluded according to the criteria listed in Table 2. Two lists of 'included' and 'excluded' articles will be collated, and reasons for exclusions will be noted. This list will be assessed by a second reviewer in order to accept or reject the 'excluded' articles. Where there is a difference in opinion, discussions will be held until a consensus is reached.

Table 2. Inclusion and exclusion criteria

	Inclusion criteria	Exclusion criteria
Subject of study	- Human campylobacteriosis	- Animal studies - Mixed pathogen studies - Non-GI mode of transmission
Method of study	- Meta-analysis and systematic reviews - Experimental studies - Observational studies including cohort studies, case-control studies and case series	- Modelling studies - Expert opinions - Narrative reviews - Letters to the editor, editorials, - Animal studies - Case studies
Characteristics of study	- No restriction on the time period - English Language	- Language other than English
Population	- Human population	- Animal population

Outcome of interest	- Incubation period of <i>campylobacter</i> spp. with or without any other outcome	- Incubation period of non-GI diseases - Period of infectiousness - Shedding of GI pathogens - Carriage

Quality assessment of identified studies

The studies to be included in this review do not necessarily aim primarily to measure the incubation time of *Campylobacter*, but rather to report an outbreak, or the outcome of an experimental study set to describe a different characteristic of the pathogen. In light of this, the quality assessment to be carried out will not aim to assess the quality of the data reported, but instead, the quality of the individual studies according to a checklist adapted from Todkill et al (Table 3).

The list of included studies (from phases 1 and 2) will be reviewed further to assess the quality and to ensure that the incubation period is recorded or can be calculated or derived from the information provided (phase 3). This phase will be carried out by both reviewers.

Table 3. Checklist for assessing individual studies reporting incubation period (adapted from Todkill et al)

Criteria	Component
Exposure	• Clearly defined exposure e.g. identification of implicating food vehicle • Exposure linked epidemiologically or microbiologically to outcome

	• Exclusion of other potential sources
Diagnosis	• Microbiological confirmation (human, food or environmental confirmation) • Specific and sensitive case definition for clinical cases • Time constraints on case definitions to exclude very early or very late cases
Accuracy of measurement	• Clearly defined exposure time (point source or continuous exposure) • Reliability of onset times considering method and delay of data collection during epidemiological investigation • Accuracy of reported onset time (hourly, 6-hourly, daily)
Ascertainment of bias	• Identification of exposed group and reporting of onset on all or part of exposed group • Exclusion of background cases • Exclusion of secondary cases and person to person transmission

Data extraction

Once the selection phase has been completed, full texts of all selected studies will be reviewed by the main reviewer to collect the required information. In addition to collecting

data on the summary statistics reported in the studies, the actual time of onset following exposure will be extracted from epidemic curves and summary tables where available.

Table 4. Data to be extracted from studies

Section	Information to be collected
General information	- Year of publication - Title of article - Authors - Type of publication (journals, conference abstract, grey literature, etc.) - PubMed ID (where applicable)
Study characteristics	- Year of study - Study design (cohort, case-control, experimental, case series) - Country of study - Age distribution - Comments on method or quality of study
Organism characteristics	- Infectious agent - Species - Subtype
Outcome data/ results	- Case definition - Reported incubation period (mean, median mode and range) - Derived or calculated incubation period (raw data extracted) - Source of calculated data (epicurve or author description)

Other outcome data	- Incubation period to particular symptoms
Factors that could affect incubation time	- No. of people affected / Total no. exposed (attack rate) - Setting - Mode of transmission - Food vehicle (for foodborne infections only)
Any other relevant information	- Any other relevant information

Data synthesis

The characteristics of the included studies will be summarised and a descriptive analysis of the actual and summarised data will be conducted. In order to assess homogeneity of the incubation period reported, the distribution of the reported onset times will be compared.

Statistical analysis will be conducted to confirm the similarities or differences in the distribution patterns of the reported incubation periods. I^2 value will be calculated and Kolmogorov-Smirnov (KS test) analysis will be conducted to test for heterogeneity between the studies and if any observed difference is due to chance. Hierarchical cluster analysis will be employed to describe the pattern of heterogeneity.

Regression models will be applied to identify and explain the factors that might influence incubation period according to the study characteristics.

Dissemination plans

The systematic review will be submitted to a peer reviewed journal for publication and will also contribute to a thesis chapter in a PhD project.

APPENDIX 2: PROTOCOL FOR COLLECTION OF INDIVIDUAL PATIENT DATA

Overview

The national surveillance system for general outbreaks in England began in 1992 and evolved into the PHE electronic Foodborne and non-Foodborne Gastrointestinal Outbreak Surveillance System (e-FOSS) in 2009. It collates information of foodborne and non-foodborne outbreaks of Shiga-toxin producing *Escherichia coli* O157 (STEC O157) including person to person transmission. Although substantial, the information gathered by this surveillance system is incomplete, for example, outbreaks of non-O157 STEC are not reported. The system is also restricted to summaries of outbreak findings and does not archive study datasets.

Routine enhanced surveillance of STEC has been ongoing in England and Wales since 1 January 2009 using the National Enhanced Surveillance System (NESS). The local laboratories report presumptive isolates to Public Health England Centres (PHEC); and through completion of standardized questionnaires, the surveillance system collates epidemiologic information of all confirmed STEC cases. This database contains a substantial amount of information on outbreak associated cases, and when used in conjunction with the e-FOSS database, a considerable amount of information is available on the outbreaks. This piece of work aims to collate individual patient data (raw data) of STEC outbreaks reported and investigated in England between 2009 and 2016.

Identification of outbreaks

Published and unpublished investigated outbreaks of STEC will be compiled. Outbreaks will be identified through eFOSS and NESS.

Data extraction and collation

Records of outbreaks will be reviewed with the purpose of identifying individual patient data.

Details to be extracted include but not limited to: demographic information (name, age, gender) infecting pathogen, date of onset, date of exposure, duration of illness, reported symptoms treatment, hospitalisation, underlying conditions and date of interview. Analytical study datasets and incident reports will also be sought to retrieve necessary information.

The information collected will be available in a format that enables easy accessibility by the end user with an interface for querying and reporting.

It is anticipated that patient details will be available and obtained in various formats, therefore the individual participant data will be cleaned and inputted using a consistent data format. The extracted data will be cleaned to remove duplicate records and inconsistencies, then pooled and stored in a purpose-built database.

Data sources

eFOSS: Outbreaks of STEC O157 reported to PHE's electronic foodborne and non-foodborne gastrointestinal outbreak surveillance system (eFOSS) between the study period.

NESS: Individual patient data of cases associated with STEC O157 and non-STEC O157 outbreaks reported to the PHE STEC National Surveillance System.

Information governance

The information collected will only be used for the purpose of analysing the data and understanding the distribution of incubation period of STEC. Patient identifiable data collected will be kept to a minimum and stored in line with the Caldicott guidelines. Data will be stored in a password protected database.

Estimated timeline

Task	Potential deadline
Protocol written and finalised	November 2015

Protocol reviewed and approved by PHE	June 2016
Visiting worker agreement approved by PHE	November 2016
Data collection for study commences	January 2017
Database for individual patient data created	March 2017

APPENDIX 3: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF *CAMPYLOBACTER*

Title	Reason for exclusion
Blaser et al. <i>Campylobacter</i> enteritis associated with foodborne transmission. <i>American Journal of Epidemiology</i> 1982; 116:886 – 894.	Non-point source outbreak. Two possible outbreaks overlapping.
Blaser et al. Outbreaks of <i>Campylobacter</i> enteritis in two extended families: evidence for person-to-person transmission. <i>The Journal of Paediatrics</i> 1981; 98:254-257.	Non-point source outbreak, secondary transmission evident
Braeye et al. A large community outbreak of gastroenteritis associated with consumption of drinking water contaminated by river water, Belgium, 2010. <i>Epidemiology and Infection</i> 2015; 143:711-719.	Difficult to distinguish between primary and secondary cases
Centre for Disease Prevention and Control. Multistate outbreak of <i>Campylobacter jejuni</i> infections associated with undercooked chicken livers — Northeastern United States, 2012. <i>Morbidity and mortality weekly report</i> 2013; 62:874-875.	Incubation period not reported
Centre for Disease Prevention and Control. Outbreak of <i>Campylobacter jejuni</i> infections associated with drinking unpasteurized milk procured through a cow-leasing program --- Wisconsin, 2001. <i>Morbidity and mortality weekly report</i> 2002; 51:548-549.	Exposure time not clearly defined

de Perio MA et al. <i>Campylobacter</i> infection in poultry-processing workers, Virginia, USA, 2008–2011. <i>Emerging Infectious Diseases</i> 2013; 19:286-288.	Prolonged exposure
DeFraités RF et al. An outbreak of <i>Campylobacter</i> enteritis associated with a community water supply on a U.S. military installation. <i>Medical Surveillance Medical Report</i> 2014; 21:10-15.	Continuous exposure
Deming MS et al. <i>Campylobacter</i> enteritis at a university: transmission from eating chicken and from cats. <i>American Journal of Epidemiology</i> 1987; 126:526-534.	Exposure time not clearly defined
Engberg J et al. Water-borne <i>Campylobacter jejuni</i> infection in a Danish town-a 6-week continuous source outbreak. Clinical Microbiology and Infection 1998; 4:648-656.	Continuous exposure
Fahey et al. An outbreak of <i>Campylobacter jejuni</i> enteritis associated with failed milk pasteurisation. <i>Journal of Infection</i> 1995; 31:137-143.	Exposure time not clearly defined
Finch MJ and Blake PA. Foodborne outbreaks of Campylobacteriosis: The United States experience. <i>American Journal of Epidemiology</i> 1985; 122:262-268.	Review of several outbreaks some of which are already included in the review
Gardner TJ et al. Outbreak of campylobacteriosis associated with consumption of raw peas. <i>Clinical Infectious Diseases</i> 2011; 53:26-32.	Extended exposure period
Graham C et al. Outbreak of campylobacteriosis following pre-cooked sausage consumption. <i>Australian and New Zealand Journal of Public Health</i> 2005; 29:507-510.	Non-point source outbreak. Exposure happened over 2 days

Gubbels S et al. A waterborne outbreak with a single clone of <i>Campylobacter jejuni</i> in the Danish town of Køge in May 2010. <i>Scandinavian Journal of Infectious Diseases</i> 2012; 44:586-594.	Continuous exposure
Gunnarsson H and Swedhem Å. The usefulness of Diffusion-In-Gel-ELISA in clinical practice as illustrated by a <i>Campylobacter jejuni</i> outbreak. <i>Journal of Immunological Methods</i> 1998; 215:135-144.	Not much information to confirm if outbreak is point source
Hennessy EP. An outbreak of campylobacteriosis amongst directing staff and students at the infantry training centre, Brecon, Wales, March 2004. <i>Journal of the Royal Army Medical Corps</i> 2004; 150:175-178.	Exposure time not clearly defined
Horn BJ and Lake RJ. Incubation period for campylobacteriosis and its importance in the estimation of incidence related to travel. <i>Euro Surveillance: European Communicable Disease Bulletin</i> 2013; 18.	Incubation period reported for proportion of cases not individual cases
Jakopenac I et al. A large waterborne outbreak of campylobacteriosis in Norway: the need to focus on distribution system safety. <i>BMC Infectious Diseases</i> 2008; 8:128.	Exposure time not clearly defined
Kuusi M et al. A large outbreak of campylobacteriosis associated with a municipal water supply in Finland. <i>Epidemiology and Infection</i> 2005; 133:593-601.	Exposure time not clearly defined

McNaughton RD et al. Outbreak of <i>Campylobacter</i> enteritis due to consumption of raw milk. <i>Canadian Medical Association Journal</i> 1982; 126:657.	Exposure time not clearly defined
Mentzing L. Waterborne outbreaks of <i>Campylobacter</i> enteritis in Central Sweden. <i>The Lancet</i> 1981; 318:352-354.	Exposure time not clearly defined
Møller-Stray J et al. Two outbreaks of diarrhoea in nurseries in Norway after farm visits, April to May 2009. <i>Euro Surveillance: European Communicable Disease Bulletin</i> 2012; 17.	Incubation period not reported
Morgan D et al. An outbreak of <i>Campylobacter</i> infection associated with the consumption of unpasteurised milk at a large festival in England. <i>European Journal of Epidemiology</i> 1994; 10:581-585.	Exposure time not clearly defined. Exposure possibly occurred over 3 days at a festival
O'Leary MC et al. A continuous common-source outbreak of campylobacteriosis associated with changes to the preparation of chicken liver pâté. <i>Epidemiology and Infection</i> 2009; 137:383-388.	Exposure time not clearly defined. Exposure occurred at intervals when cases dined
Porter IA and Reid TM. A milk-borne outbreak of <i>Campylobacter</i> infection. <i>The Journal of Hygiene</i> 1980; 84:415.	Exposure time not clearly defined. Date of exposure unknown, as raw milk was distributed on a certain day but day of actual consumption not recorded.

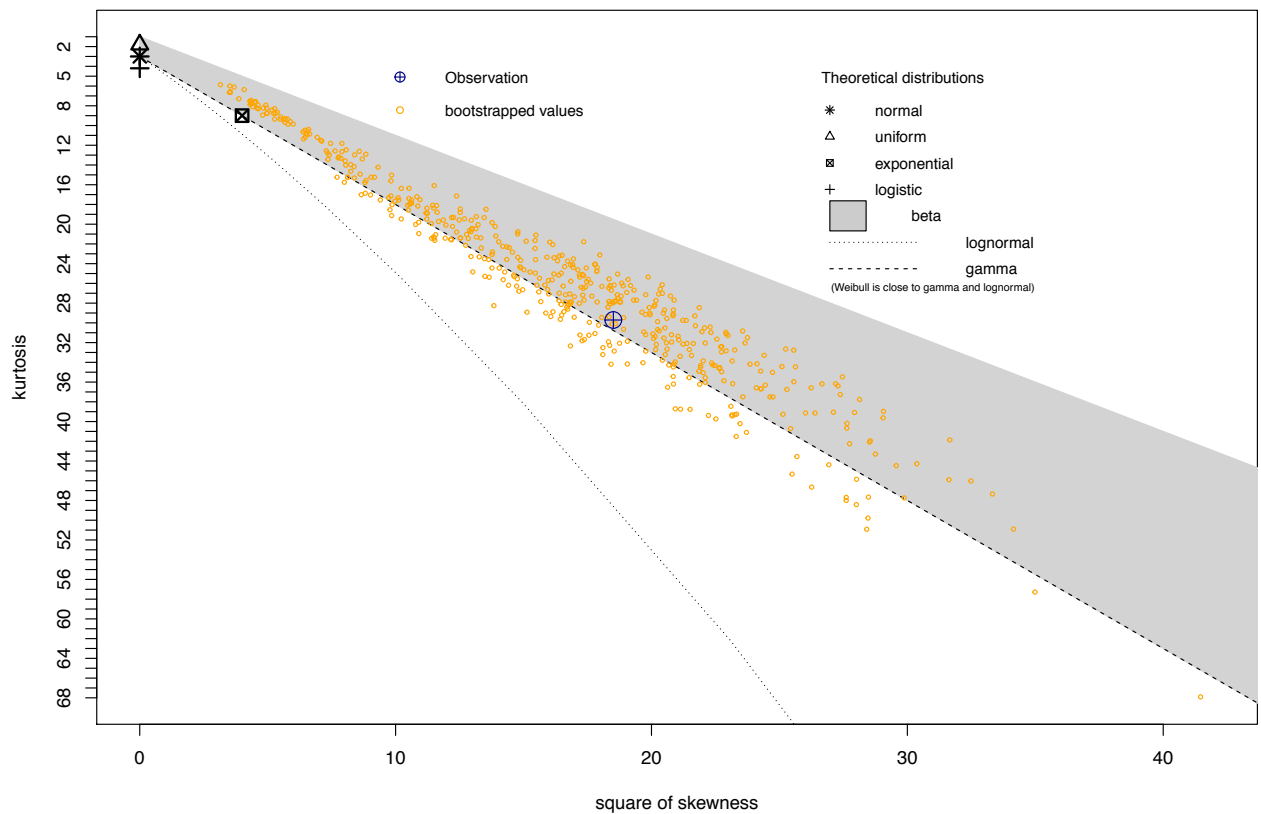
Potter ME et al. Human <i>Campylobacter</i> infection associated with certified raw milk. <i>American Journal of Epidemiology</i> 1983; 11:475-483.	Incubation period not reported
Riordan T et al. A point source outbreak of <i>Campylobacter</i> infection related to bird-pecked milk. <i>Epidemiology and Infection</i> ; 110:261.	Exposure time not clearly defined
Rogol M et al. Waterborne outbreak of <i>Campylobacter</i> enteritis. <i>European Journal of Clinical Microbiology</i> 1983; 2:588-590.	Exposure time not clearly defined
Sacks JJ et al. Epidemic campylobacteriosis associated with a community water supply. <i>American Journal of Public Health</i> 1986; 76:424-428.	Exposure time not clearly defined
Taylor DN et al. Waterborne transmission of <i>Campylobacter</i> enteritis. <i>Microbial Ecology</i> 1982; 8:347-354.	Incubation period not reported
Tettmar RE and Thornton EJ. An outbreak of <i>Campylobacter</i> enteritis affecting an operational Royal Air Force unit. <i>Public Health</i> 1981; 95:69-73.	Exposure time not clearly defined
Unicomb LE et al. Outbreaks of Campylobacteriosis in Australia, 2001 to 2006. <i>Foodborne Pathogens and Disease</i> 2009; 6:1241-1250.	Not point source outbreak
Vierikiko A et al. Domestically acquired <i>Campylobacter</i> infections in Finland. <i>Emerging Infectious Diseases</i> 2004; 10:127-130.	Incubation period not reported

Wood RC et al. <i>Campylobacter</i> enteritis outbreaks associated with drinking raw milk during youth activities: A 10-year review of outbreaks in the United States. <i>Journal of the American Medical Association</i> 1992; 268:3228-3230.	Not point source outbreak
Yanagisawa S. Large outbreak of <i>Campylobacter</i> enteritis among school children. <i>The Lancet</i> 1980; 316:153.	Incubation period not reported
Zeiger M et al. Outbreak of campylobacteriosis associated with a long-distance obstacle adventure race--Nevada, October 2012. <i>Morbidity and mortality weekly report</i> 2014; 63:375-378.	Exposure time not clearly defined. Exposure was an outdoor event that took place over two days.

**APPENDIX 4: DETAILS OF STUDIES INCLUDED AND DATA
EXTRACTED IN THE SYSTEMATIC REVIEW AND META-
REGRESSION OF THE INCUBATION PERIOD OF
CAMPYLOBACTERIOSIS**

Please see accompanying USB drive

APPENDIX 5: CULLEN AND FREY PLOT SHOWING POTENTIAL DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF *CAMPYLOBACTER*



APPENDIX 6: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF STEC

Article	Reason for exclusion
Ackman D, Marks S, Mack P, Caldwell M, Root T, Birkhead G. Swimming-associated haemorrhagic colitis due to <i>Escherichia coli</i> O157:H7 infection: evidence of prolonged contamination of a fresh water lake. Epidemiol Infect. 1997;119:1-8.	Not point source exposure. Park visits occurred over several days and difficult to pin point exact onset dates of all cases.
Aldabe B, Delmas Y, Gault G, Vendrely B, Llanas B, Charron M, et al. Household transmission of haemolytic uraemic syndrome associated with <i>Escherichia coli</i> O104:H4, south-western France, June 2011. Euro Surveill. 2011;16(31)	Person to person transmission in the household. Unsure of exposure times of secondary cases
Aragon T, Garcia B. Outbreak of Shiga Toxin-Producing <i>Escherichia coli</i> O157 Infection Associated with a Restaurant. Population Health Division, San Francisco Department of Public Health 2013.	Exposure occurred on two possible days. And although the authors reported median and range IP, they did not specify when each case was exposed, so IP not definite.

Cody S, Glynn K, Farrar J, Cairns K, Griffin P, Kobayashi J, et al. An Outbreak of <i>Escherichia coli</i> O157:H7 Infection from Unpasteurized Commercial Apple Juice. Ann Intern Med. 1999; 130:202-209.	Commercially distributed products. A subset of cases had specific exposure dates, however the onset date of these cases was not specified.
Devaux I, Varela-Santos C, Payne-Hallstrom L, Takkinen J, Bogaardt C, Coulombier D. Investigation of travel-related cases in a multinational outbreak: example of the Shiga-toxin producing <i>E. coli</i> outbreak in Germany, May–June 2011. Epidemiol Infect. 2015	travel related cases. Specific dates of exposure unknown so estimated from cases that spent 3 days in Germany and calculated as the mean date of stay to the onset.
Duncan L, Mal V, Carter A, Carison J, Borczyk A, Karmati M. Outbreak of gastrointestinal disease – Ontario. Can Dis Wkly Rep. 1987;13(2):5-8.	Exposure occurred over two days and authors did not specify which case was exposed when even though they provided date of onset

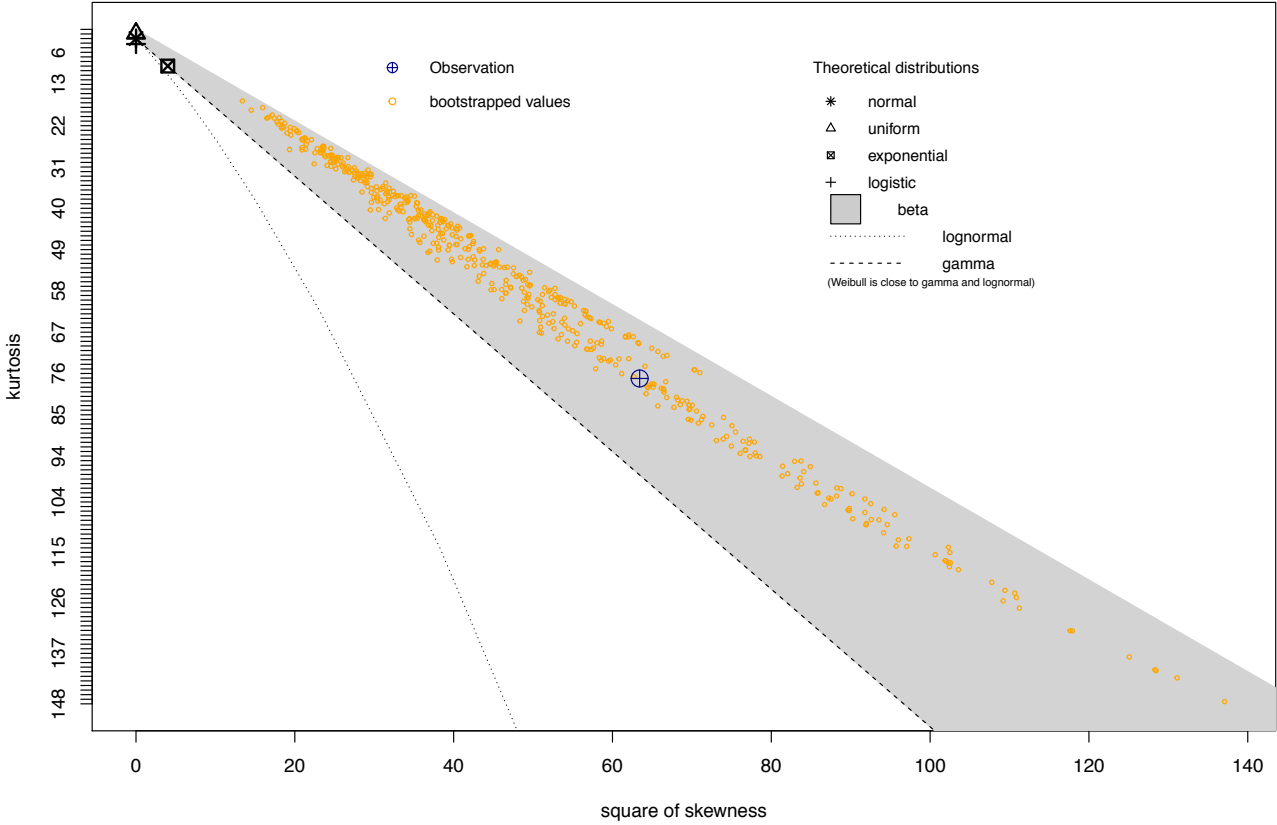
Ferguson D, Scheftel J, Cronquist A, Smith K, Woo-Ming A, Anderson E, et al. Temporally distinct <i>Escherichia coli</i> O157 outbreaks associated with alfalfa sprouts linked to a common seed source – Colorado and Minnesota, 2003. Epidemiol Infect. 2005;133:439-447.	Non-point source. Cases were exposed at different times with no individual exposure date reported
Frank C, Werber D, Cramer J, Askar M, Faber M, an der Heiden M, et al. Epidemic Profile of Shiga-Toxin–Producing <i>Escherichia coli</i> O104:H4 Outbreak in Germany. N Engl J Med 2011;365:1771-1780.	No exposure dates reported
Robert Koch Institute. Report: Final presentation and evaluation of epidemiological findings in the EHEC O104:H4 outbreak, Germany 2011. Berlin 2011.	Exposure dates unknown and incubation period estimated as a probability
Riley L, Remis R, Helgerson S, McGee H, Wells J, Davis B, et al. Hemorrhagic colitis associated with a rare <i>Escherichia coli</i> serotype. N Engl J Med. 1983;308:681-685.	Not point source

Sharapov U, Wendel A, Davis J, Keene W, Farrar J, Sodha S, et al. Multistate Outbreak of <i>Escherichia coli</i> O157:H7 Infections Associated with Consumption of Fresh Spinach: United States, 2006. J Food Prot. 2016;79(12):2024-2030	Not point source. Widely distributed product.
Sonpal, Niket, et al. (2011) "A Salad a Day won't Keep <i>E. coli</i> Away: A Spectrum Case Series of <i>E. coli</i> O157: H7 in Vegetarians." <i>American Journal of Gastroenterology</i> . Conference: 76th Annual Scientific Meeting of the American College of Gastroenterology. Washington, DC United States.	Case series without exposure times
Werber D, King L, Muller L, Follin P, Buchholz U, Bernard H, et al. Associations of Age and Sex With the Clinical Outcome and Incubation Period of Shiga toxin-producing <i>Escherichia coli</i> O104:H4 Infections, 2011. Am J Epidemiol. 2013;178(6):984-992.	Unsure about the IP reported. Reviewing the referenced paper for cohort 1, I could not identify the IP reported. Furthermore, the authors stated that if multiple exposure was possible, a median of exposure days was taken.
Yahata Y, Misaki T, Ishida Y, Nagira M, Watahiki M, et al. Epidemiological analysis of a large enterohaemorrhagic <i>Escherichia coli</i> O111 outbreak in Japan associated with haemolytic uraemic syndrome and acute encephalopathy. Epidemiol Infect. 2015;143:2721-2732.	Different branches of a restaurant and different exposure dates

APPENDIX 7: DETAILS OF STUDIES INCLUDED AND DATA EXTRACTED IN THE SYSTEMATIC REVIEW AND META-REGRESSION OF THE INCUBATION PERIOD OF STEC

Please see accompanying USB drive

APPENDIX 8: CULLEN AND FREY PLOT SHOWING POTENTIAL
DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF STEC



APPENDIX 9: STEC ENHANCED SURVEILLANCE QUESTIONNAIRE

Shiga toxin-producing *Escherichia coli* Enhanced surveillance questionnaire

SECTION A: QUESTIONNAIRE DETAILS

Interviewer name: Interview date :
Interviewer office: Interviewer Telephone:
Person interviewed name:
Details from: if other specify

SECTION B: CASE CLASSIFICATION – SEE GUIDANCE NOTES FOR DEFINITIONS

Is this case: ☐ Asymptomatic
If secondary, name primary case:
Investigation is:
Outbreak keyword or number:
Outcome: Recovered HUS/TTP if died, then → Date of death:
Select all that apply Still ill Died

SECTION C: PERSONAL DETAILS

First name: Family name:
Address:
Postcode: Tel (h): Tel (m):
Email:
Sex: Date of birth (dd/mm/yyyy): Age: yrs
NHS No: GP name:
GP address: GP Tel:
Are there any children living in the household? (other than the case) Y ☐ N ☐ U ☐
Occupation: Tick if any of the below risk groups apply
☐ Foodhandler (e.g. handle food professionally) ☐ Work in/attend healthcare setting
☐ Work in/attend childcare setting ☐ Work in contact with faeces (e.g. lab, farm etc.)
☐ Have difficulty maintaining personal hygiene ☐ Other risk category
If yes to any of the above, details:
Work/School: Tel:
Address:
Postcode: Last date:
Ethnicity:

If the case is asymptomatic, please send form to your HPU

SECTION D: SYMPTOMS OF ILLNESS

Onset date:	Still ill:	If N, then →			Duration of illness (d):	
Symptoms experienced	Yes	No	Not Sure	Ongoing	Duration (d)	Date of onset
Diarrhoea (3 or more loose stools in 24hrs)	☐	☐	☐	☐		
Bloody stools	☐	☐	☐	☐		
Nausea	☐	☐	☐	☐		
Vomiting	☐	☐	☐	☐		
Abdominal pain	☐	☐	☐	☐		
Fever	☐	☐	☐	☐		

Other (specify):

Sought healthcare: ☐ NHS Direct ☐ GP visit ☐ A&E ☐ Other (specify):

Submitted stool sample:

Admitted to hospital for this illness:

Admission date:

Hospital name:

Duration of stay
(d):

Self/Medicated with antibiotics:

If Y, specify:

Self/Medicated with antidiarrhoeals:

If Y, specify:

Detail any other health concerns:

SECTION E: TRAVEL IN THE WEEK PRIOR TO ILLNESS

Travelled **OUTSIDE** of the UK:

Specify countries visited (from most recent to least recent)

Country/Region	Date arrived	Date departed	Details

Travelled **WITHIN** the UK:

If yes, Specify towns/resorts visited (from most recent to least recent)

Town/Resort	Date arrived	Date departed	Details

If yes, describe the type of accommodation used:

Name of accommodation:

SECTION F: FOOD HISTORY IN THE WEEK PRIOR TO ILLNESS

Ate out:

Venues	Yes	No	Name/Location of place	Dates:	Eaten:
Restaurant 1	☐	☐			
Restaurant 2	☐	☐			
Takeaway1	☐	☐			
Takeaway2	☐	☐			
Café/Canteen	☐	☐			
Party/BBQ/function	☐	☐			
Other1	☐	☐			
Other2	☐	☐			

Do you handle/prepare most of the food within the household: Always Mostly Occasionally Never

Ate or handled any of the following:

	Yes	No	Product (e.g. boneless, skinless chicken breast)	Where purchased (e.g. Asda Acton, London)
HANDLER	☐	☐	Raw beef	
	☐	☐	Raw poultry	
	☐	☐	Raw lamb	
	☐	☐	Raw pork/gammon	
	☐	☐	Other raw meat (e.g. game, goat, ostrich etc.)	
	☐	☐	Raw vegetables	
	☐	☐	Pet/animal feed	
	Any meat ☐ ☐ <i>If yes, indicate below:</i>			
CONSUMER	☐	☐	Cooked Beef	
	☐	☐	Cooked poultry	
	☐	☐	Cooked lamb	
	☐	☐	Cooked pork/gammon	
	☐	☐	Other cooked meat (e.g. game, goat, ostrich etc.)	
	☐	☐	Cured meats	
	☐	☐	Other processed meat	
	☐	☐	Fish	
☐	☐	Shellfish		
☐	☐	Pasteurised milk		

CONSUMER	Unpasteurised milk	☐	☐
	Hard cheese	☐	☐
	Soft cheese	☐	☐
	Yoghurt/ Fromage Frais	☐	☐
	Cream	☐	☐
	Ice Cream	☐	☐
	Unpasteurised dairy products	☐	☐
	Pre-packaged salad	☐	☐
	Other salad	☐	☐
	Raw vegetables	☐	☐
	Raw fruit	☐	☐
	Soft fruit/berries	☐	☐
	Pre cut fruits	☐	☐
	Other Raw fruit	☐	☐
	Sprouted seeds/ beansprouts	☐	☐
	Fresh herbs	☐	☐
	Fruit juices	☐	☐
	Pre-packaged sandwiches etc.	☐	☐
	Other foods (e.g.nuts, confectionery, etc.)	☐	☐

SECTION G: WATER EXPOSURE IN THE WEEK PRIOR TO ILLNESS

Drank unboiled water from any of the following:

Water supply	Yes	No	Details
Mains (municipal) water	☐	☐	
Private water supply (spring/well/borehole)	☐	☐	
Bottled water	☐	☐	
Unboiled river/stream/lake water	☐	☐	
Exposed to floodwater:	Yes ☐	No ☐	Not sure ☐

If yes, details:

Experienced any household drainage/plumbing problems: Yes ☐ No ☐ Not sure ☐

If yes, details:

Participated in any of the following activities – either recreationally or for occupation:

Activity	Fresh water	Sea Water	No	Details

Swimming/Paddling ☐ ☐ ☐
 Other (e.g. canoeing, ☐ ☐ ☐
 fishing, sailing, ☐
 surfing). ☐

Is it possible that water was accidentally swallowed during any of the above:

If yes, details:

SECTION H: ANIMAL CONTACT IN THE WEEK PRIOR TO ILLNESS

Contact with domestic animals/pets: Yes ☐ No ☐

If yes, indicate which animals below:

Dogs ☐ Cats ☐ Rabbits ☐ Rodents ☐ Reptiles ☐ Birds ☐ Fish ☐ Other: ☐

Did any of the above animals have diarrhoea: Yes ☐ No ☐ Not sure ☐

If yes, specify:

Was there a veterinary investigation? Yes ☐ No ☐ Not sure ☐

Contact with non domestic animals: Yes ☐ No ☐ Not sure ☐

Cattle ☐ Calves ☐ Sheep ☐ Lambs ☐ Goats ☐ Horses ☐ Pigs ☐ Reptiles ☐

Poultry ☐ Rabbits ☐ Deer ☐ Birds ☐ Fish ☐ Rodents ☐ Other: ☐

Lives on, works at or access to a private farm: Yes ☐ No ☐ Not sure ☐

Attended an agricultural event e.g horse show Yes ☐ No ☐ Attended date:
 Not sure ☐

Visited a farm/petting zoo/bird reserve or such: Yes ☐ No ☐ Visit date:
 Not sure ☐

If yes, specify:

If yes, Handled the animals: Yes ☐ No ☐ Not sure ☐

If yes, specify:

Bottle fed any of the animals: Yes ☐ No ☐ Not sure ☐

Washed hands before eating food or before leaving: Yes ☐ No ☐ Not sure ☐

If yes, specify:

Consumed any food whilst there:

If yes, specify:

Was the food (tick all that apply):

☐ Purchased on premises OR ☐ Brought from home/elsewhere
☐ Eaten in a separate area OR ☐ Eaten whilst in contact with the animals
☐ Eaten at a picnic table OR ☐ Eaten while sat on the grass/soil

SECTION I: ENVIRONMENTAL EXPOSURE IN THE WEEK PRIOR TO ILLNESS

Walked in a paddock/field where farm animals graze: Yes ☐ No ☐ Not sure ☐

If yes, specify:

Taken any day trips (beach, countryside etc.): Yes ☐ No ☐ Not sure ☐

If yes, specify:

In contact with wildlife species or their droppings: Yes ☐ No ☐ Not sure ☐

If yes, specify:

Had contact with soil, manure or sewage:

Yes ☐

No ☐

Not sure ☐

If yes, specify:

SECTION J: ANY OTHER COMMENTS RELEVANT TO THIS CASE

Can the case be contacted again if further details are required: ☐

PHE Centre Staff: Please email this form to vtec@phe.gov.uk

APPENDIX 10: LIST OF STUDIES EXCLUDED FROM REVIEW AND REASON FOR EXCLUSION IN THE SYSTEMATIC REVIEW OF *SALMONELLA* TYPHI

Article	Reason for exclusion
Agarwal V, Jalgaonkar PD, Pathak AA, Saoji AM. An outbreak of multidrug resistant typhoid fever in Nagpur. J Assoc Physicians India. 1992 Jun;40(6):416.	No incubation period reported
Al-Sanouri TM, Paglietti B, Haddadin A, Murgia M, Bacciu D, Youssef M, et al. Emergence of plasmid-mediated multidrug resistance in epidemic and non-epidemic strains of <i>Salmonella enterica</i> serotype Typhi from Jordan. J Infect Dev Ctries. 2008 Aug 30;2(4):295–301.	Study on microbiology of S. Typhi with no report on incubation period
al-Zubaidy AA, el Bushra HE, Mawlawi MY. An outbreak of typhoid fever among children who attended a potluck dinner at Al-Mudhnab, Saudi Arabia. East Afr Med J. 1995 Jun;72(6):373–5.	Exposure could have occurred over two days or one of two days
Anderson ES, Hobbs BC. Studies of the strain of <i>Salmonella</i> Typhi responsible for the Aberdeen typhoid outbreak. Isr J Med Sci. 1973 Feb;9(2):162–74.	Study on microbiology of S. Typhi with no report on incubation period
Arora RK, Gupta A, Joshi NM, Kataria VK, Lall P, Anand AC. Multidrug resistant typhoid fever: study of an outbreak in Calcutta. Indian Pediatr. 1992 Jan;29(1):61–6.	Study on microbiology of S. Typhi with no report on incubation period

Arya SC, Agarwal N. Re: outbreak of typhoid fever in vaccinated members of the French Armed Forces in the Ivory Coast. Eur J Epidemiol. 2006;21(5):407.	Microbiological focus of a study included in the review
Ash I, Mckendrick GD, Robertson MH, Hughes HL. Outbreak of typhoid fever connected with corned beef. Br Med J. 1964 Jun 6;1(5396):1474–8.	Non-point source outbreak. Corned beef purchased on a particular day, but unsure of date of consumption.
Aye TT, Siriarayapon P. Typhoid fever outbreak in Madaya Township, Mandalay Division, Myanmar, September 2000. J Med Assoc Thai. 2004 Apr;87(4):395–9.	Non-point source outbreak.
Ayliffe GA, Geddes AM, Pearson JE, Williams TC. Spread of <i>Salmonella</i> Typhi in a maternity hospital. J Hyg (Lond). 1979 Jun;82(3):353–9.	Source of exposure and mode of transmission unknown. No incubation period reported.
Aypak A, Celik AK, Aypak C, Cikman O. Multidrug resistant typhoid fever outbreak in Ercek Village-Van, Eastern Anatolia, Turkey: clinical profile, sensitivity patterns and response to antimicrobials. Trop Doct. 2010 Jul;40(3):160–2.	No incubation period reported
Ayyagari A, Pal N. Outbreak of typhoid fever due to multi-resistant <i>Salmonella</i> Typhi in northern India--a preliminary report. Trans R Soc Trop Med Hyg. 1991 Apr;85(2):302.	No incubation period reported

Bahrmand AR, Velayati AA. Antimicrobial resistance pattern and plasmid profile of <i>Salmonella</i> Typhi isolated from an outbreak in Tehran province. Scand J Infect Dis. 1997;29(3):265–9.	Study on microbiology of S. Typhi with no report on incubation period
Bayram Y, Güdücüoğlu H, Otlu B, Aypak C, Gürsoy NC, Uluç H, et al. Epidemiological characteristics and molecular typing of <i>Salmonella enterica</i> serovar Typhi during a waterborne outbreak in Eastern Anatolia. Ann Trop Med Parasitol. 2011	Non-point source outbreak.
Bhunja R, Hutin Y, Ramakrishnan R, Pal N, Sen T, Murhekar M. A typhoid fever outbreak in a slum of South Dumdum municipality, West Bengal, India, 2007: evidence for foodborne and waterborne transmission. BMC Public Health. 2009 Apr 27;9:115.	Non-point source outbreak and possible secondary transmission.
Bissett ML, Powers C, Wood RM. Immunofluorescent identification of <i>Salmonella</i> Typhi during a typhoid outbreak. Appl Microbiol. 1969 Apr;17(4):507–11.	Study on microbiology of S. Typhi with no report on incubation period
Blum LS, Dentz H, Chingoli F, Chilima B, Warne T, Lee C, et al. Formative investigation of acceptability of typhoid vaccine during a typhoid fever outbreak in Neno District, Malawi. Am J Trop Med Hyg. 2014 Oct;91(4):729–37.	Qualitative study on vaccine acceptability. No incubation period reported.

Bradarić N, Punda-Polić V, Milas I, Ivić I, Grgić D, Radosević N, et al. Two outbreaks of typhoid fever related to the war in Bosnia and Herzegovina. Eur J Epidemiol. 1996 Aug;12(4):409–12.	Unable to determine time of exposure
Bradley WH, Evans LW, Taylor I. A hospital outbreak of typhoid fever. J Hyg (Lond). 1951 Sep;49(2–3):324–34.	Exposure seemed to have occurred over 3 days
Centers for Disease Control and Prevention (CDC). Notes from the field: <i>Salmonella</i> Typhi infections associated with contaminated water--Zimbabwe, October 2011-May 2012. MMWR Morb Mortal Wkly Rep. 2012 Jun 15;61(23):435.	Non-point source outbreak.
Coovadia YM, Gathiram V, Bhamjee A, Garratt RM, Mlisana K, Pillay N, et al. An outbreak of multi-resistant <i>Salmonella</i> Typhi in South Africa. Q J Med. 1992 Feb;82(298):91–100.	Study on microbiology of <i>S. Typhi</i> with no report on incubation period
Coupee WRM, Newell KW, Payne DJH. An Outbreak of Typhoid Fever associated with Canned Ox-Tongue. Lancet. 1956;1057–9.	Contaminated product bought over two weeks, even though author claims exposure occurred on one day
Egoz N, Shihab S, Leitner L, Lucian M. An outbreak of typhoid fever due to contamination of the municipal water supply in northern Israel. Isr J Med Sci. 1988 Nov;24(11):640–3.	Non-point source outbreak. Community supplied water

Engleberg NC, Barrett TJ, Fisher H, Porter B, Hurtado E, Hughes JM. Identification of a carrier by using Vi enzyme-linked immunosorbent assay serology in an outbreak of typhoid fever on an Indian reservation. J Clin Microbiol. 1983 Dec;18(6):1320–2.	Exposure time unsure. No incubation period reported
Farooqui A, Khan A, Kazmi SU. Investigation of a community outbreak of typhoid fever associated with drinking water. BMC Public Health. 2009 Dec 20;9:476.	Non-point source outbreak.
Gallay A, Vaillant V, Bouvet P, Grimont P, Desenclos JC. How many foodborne outbreaks of <i>Salmonella</i> infection occurred in France in 1995? Application of the capture-recapture method to three surveillance systems. Am J Epidemiol. 2000 Jul 15;152(2):171–7.	Review of a surveillance system
Glynn JR, Bradley DJ. The relationship between infecting dose and severity of disease in reported outbreaks of <i>Salmonella</i> infections. Epidemiol Infect. 1992 Dec;109(3):371–88.	Dose response study. No incubation period reported
Goh KT, Teo SH, Tay L, Monteiro EH. Epidemiology and control of an outbreak of typhoid in a psychiatric institution. Epidemiol Infect. 1992 Apr;108(2):221–9.	Non-point source outbreak.

Gonzalez-Cortes A, Bessudo D, Sanchez-Leyva R, Fragoso R, Hinojosa M, Becerril P. Water-borne transmission of chloramphenicol-resistant <i>Salmonella</i> Typhi in Mexico. Lancet. 1973 Sep 15;2(7829):605–7.	Non-point source outbreak.
Gupta V, Kaur U, Singh G, Prakash C, Sharma M, Aggarwal KC. An outbreak of typhoid fever in Chandigarh, North India. Trop Geogr Med. 1986 Mar;38(1):51–4.	Non-point source outbreak.
Halder KK, Dalal BS, Ghose E, Sanyal S. Chloramphenicol resistant <i>Salmonella</i> Typhi: the cause of recent out-break of enteric fever in Calcutta. Indian J Pathol Microbiol. 1992 Jan;35(1):11–7.	No incubation period reported
Hancock-Allen J, Cronquist AB, Peden J, Adamson D, Corral N, Brown K. Notes from the field: typhoid fever outbreak associated with an asymptomatic carrier at a restaurant - Weld County, Colorado, 2015. MMWR Morb Mortal Wkly Rep. 2016 Jun 17;65(23):606–7.	Non-point source outbreak.
Hardy G, Padfield CJ, Chadwick P, Partington MW. Typhoid outbreak in Kingston, Ont: experience with high-dose oral ampicillin. Can Med Assoc J. 1977 Apr 9;116(7):761–4, 767.	Undefined exposure time
Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, Lukwesa-Musyani C, Tambatamba B, Mwaba J, et al. Genomic signature of multidrug-resistant	Study on microbiology of <i>S. Typhi</i> with no report on incubation period

<i>Salmonella enterica</i> serovar Typhi isolates related to a massive outbreak in Zambia between 2010 and 2012. J Clin Microbiol. 2015 Jan;53(1):262–72.	
Henriksen SD, Riddervold HO. An outbreak of food-poisoning due to an unusual strain of <i>Salmonella</i> Typhi. Acta Pathologica Et Microbiologica Scandinavica 1952 Jun; 93:254-258.	Possible point source outbreak, however, no human confirmation and atypical of salmonella typhi. Most likely not S. Typhi outbreak. Author also mentioned that there are no clinical typhoid cases.
Hoffman TA, Ruiz CJ, Counts GW, Sachs JM, Nitzkin JL. Waterborne typhoid fever in Dade County, Florida. Clinical and therapeutic evaluation of 105 bacteremic patients. Am J Med. 1975 Oct;59(4):481–7.	Non-point source outbreak.
Jain S, Chitnis DS, Sham A, Rathi S, Inamdar S, Rindani GJ. Outbreak of chloramphenicol resistant typhoid fever. Indian Pediatr. 1987 Mar;24(3):193–7.	Study on microbiology of S. Typhi with no report on incubation period
Jordan EO, Irons EE. The Rockford (Ill) typhoid epidemic. J Infect Dis. 1912;11:21–43.	Unknown exposure time. Difficult to distinguish primary from secondary cases
Kamili MA, Ali G, Shah MY, Rashid S, Khan S, Allaqaband GQ. Multiple drug resistant typhoid fever outbreak in Kashmir Valley. Indian J Med Sci. 1993 Jun;47(6):147–51.	Non-point source outbreak.

Kato Y, Fukayama M, Adachi T, Imamura A, Tsunoda T, Takayama N, et al. Multidrug- resistant typhoid fever outbreak in travelers returning from Bangladesh. Emerging Infect Dis. 2007 Dec;13(12):1954–5.	Non-point source outbreak.
Katz DJ, Cruz MA, Trepka MJ, Suarez JA, Fiorella PD, Hammond RM. An outbreak of typhoid Fever in Florida associated with an imported frozen fruit. J Infect Dis. 2002 Jul 15;186(2):234–9.	Non-point source outbreak.
Keddy KH, Sooka A, Ismail H, Smith AM, Weber I, Letsoalo ME, et al. Molecular epidemiological investigation of a typhoid fever outbreak in South Africa, 2005: the relationship to a previous epidemic in 1993. Epidemiol Infect. 2011 Aug;139(8):1239–45.	Study on microbiology of <i>S. Typhi</i> with no report on incubation period
King CC, Chen CJ, You SL, Chuang YC, Huang HH, Tsai WC. Community-wide epidemiological investigation of a typhoid outbreak in a rural township in Taiwan, Republic of China. Int J Epidemiol. 1989 Mar;18(1):254–60.	Non-point source outbreak.
Kulkarni R, Bhandar M, Srinivasa S. Chloramphenicol resistant <i>Salmonella</i> Typhi outbreak in a rural part of north-east Karnataka. J Commun Dis. 1994 Dec;26(4):235–6.	No incubation period reported

Lewis MD, Serichantalergs O, Pitarangsi C, Chuanak N, Mason CJ, Regmi LR, et al. Typhoid fever: a massive, single-point source, multidrug-resistant outbreak in Nepal. Clin Infect Dis. 2005 Feb 15;40(4):554–61.	Single source of exposure associated with water supply, but not point source outbreak as exposure occurred over a period of months
Lin FY, Becke JM, Groves C, Lim BP, Israel E, Becker EF, et al. Restaurant-associated outbreak of typhoid fever in Maryland: identification of carrier facilitated by measurement of serum Vi antibodies. J Clin Microbiol. 1988 Jun;26(6):1194–7.	Time of exposure not reported
Loharikar A, Newton A, Rowley P, Wheeler C, Bruno T, Barillas H, et al. Typhoid fever outbreak associated with frozen mamey pulp imported from Guatemala to the western United States, 2010. Clin Infect Dis. 2012 Jul;55(1):61–6.	Non-point source outbreak.
Mallory A, Belden EA, Brachman PS. The current status of typhoid fever in the United States and a description of an outbreak. J Infect Dis. 1969 Jun;119(6):673–6.	Non-point source outbreak. Infection probably occurred over two days
Marmion DE, Naylor GRE, Stewart IO. Second attacks of typhoid fever. J Hyg (Lond). 1953 Jun;51(2):260–7.	Primary cases not well differentiated from secondary cases
Miner JR. The Incubation Period of Typhoid Fever. The Journal of Infectious Diseases. 1922;31(3):296–301.	Review of other outbreaks already included in the study

Muti M, Gombe N, Tshimanga M, Takundwa L, Bangure D, Mungofa S, et al. Typhoid outbreak investigation in Dzivaresekwa, suburb of Harare City, Zimbabwe, 2011. Pan Afr Med J. 2014;18:309.	Non-point source outbreak.
Neil KP, Sodha SV, Lukwago L, O-Tipo S, Mikoleit M, Simington SD, et al. A large outbreak of typhoid fever associated with a high rate of intestinal perforation in Kasese District, Uganda, 2008-2009. Clin Infect Dis. 2012 Apr;54(8):1091–9.	Non-point source outbreak.
Neill WA, Martin JD, Belden EA, Trotter WY. A widespread epidemic of typhoid fever traced to a common exposure. New England Journal of Medicine. 1958 Oct 2;259(14):667–72.	Date of actual exposure unknown. Date of arrival to camp was used as a proxy to calculate incubation period.
Olsen SJ, Bleasdale SC, Magnano AR, Landrigan C, Holland BH, Tauxe RV, et al. Outbreaks of typhoid fever in the United States, 1960-99. Epidemiol Infect. 2003 Feb;130(1):13–21.	Review study. No incubation period reported
Olsen SJ, Kafoa B, Win NS, Jose M, Bibb W, Luby S, et al. Restaurant-associated outbreak of <i>Salmonella</i> Typhi in Nauru: an epidemiological and cost analysis. Epidemiol Infect. 2001 Dec;127(3):405–12.	Non-point source outbreak.
Rajeev A. Quinolone resistant typhoid outbreak in an extended joint family. J Commun Dis. 1999 Dec;31(4):263–5.	Exposure time not stated

Ramanan A, Pandit N, Yeshwanth M. Unusual complications in a multidrug resistant <i>Salmonella</i> Typhi outbreak. Indian Pediatr. 1992 Jan;29(1):118–20.	Date of onset or exposure not reported
Rao RS, Amarnath SK, Sujatha S. An outbreak of typhoid due to multidrug resistant <i>Salmonella</i> Typhi in Pondicherry. Trans R Soc Trop Med Hyg. 1992 Apr;86(2):204–5.	Review paper with no incubation period reported
Rathish KC, Chandrashekar MR, Nagesha CN. An outbreak of multidrug resistant typhoid fever in Bangalore. Indian J Pediatr. 1995 Aug;62(4):445–8.	Review paper with no incubation period reported
Reller ME, Olsen SJ, Kressel AB, Moon TD, Kubota KA, Adcock MP, et al. Sexual transmission of typhoid fever: a multistate outbreak among men who have sex with men. Clin Infect Dis. 2003 Jul 1;37(1):141–4.	Non-GI transmission
Roy JS, Saikia L, Medhi M, Tassa D. Epidemiological investigation of an outbreak of typhoid fever in Jorhat town of Assam, India. Indian J Med Res. 2016 Oct;144(4):592–6.	Non-point source outbreak. Source of exposure is community water supply
Sejvar J, Lutterloh E, Naiene J, Likaka A, Manda R, Nygren B, et al. Neurologic manifestations associated with an outbreak of typhoid fever, Malawi--Mozambique, 2009: an epidemiologic investigation. PLoS ONE. 2012;7(12):e46099.	No incubation period reported

Singla N, Bansal N, Gupta V, Chander J. Outbreak of Salmonella Typhi enteric fever in sub-urban area of North India: a public health perspective. Asian Pac J Trop Med. 2013 Feb;6(2):167–8.	Non-point source outbreak and no incubation period reported
Stanwell-Smith RE, Ward LR. An international point source outbreak of typhoid fever: a European collaborative investigation. Bull World Health Organ. 1986;64(2):271–8.	No incubation period reported
Styliads S, Borczyk A. Typhoid outbreak associated with consumption of raw shellfish-- Ontario. Can Commun Dis Rep. 1994 Apr 30;20(8):63–5.	Non-point source outbreak. Food vehicle was purchased over several days
Swaddiwudhipong W, Kanlayanaphotporn J. A common-source water-borne outbreak of multidrug-resistant typhoid fever in a rural Thai community. J Med Assoc Thai. 2001 Nov;84(11):1513–7.	Non-point source outbreak. Source of exposure is community water supply
Taylor A, Santiago A, Gonzalez-Cortes A, Gangarosa EJ. Outbreak of typhoid fever in Trinidad in 1971 traced to a commercial ice cream product. Am J Epidemiol. 1974 Aug;100(2):150–7.	Non-point source outbreak. Dates ice creams were purchased or consumed is unknown. Only date of distribution is reported
Taylor JP, Shandera WX, Betz TG, Schraitle K, Chaffee L, Lopez L, et al. Typhoid fever in San Antonio, Texas: an outbreak traced to a continuing source. J Infect Dis. 1984 Apr;149(4):553–7.	Non-point source outbreak.

Torin DE, Betts SL, McClenahan JB, Phillips GM, Shikamura MT. A typhoid fever outbreak on a university campus. Arch Intern Med. 1969 Nov;124(5):606–10.	Unknown exposure, so difficult to know when cases were exposed
Usera MA, Aladueña A, Echeita A, Amor E, Gomez-Garcés JL, Ibañez C, et al. Investigation of an outbreak of Salmonella Typhi in a public school in Madrid. Eur J Epidemiol. 1993 May;9(3):251–4.	No specific onset dates reported
Weaver LT, Eccles JB, Eccles MP, Foord F. <i>Salmonella</i> Typhi infection associated with a school feeding programme. J Trop Pediatr. 1989 Dec;35(6):331–2.	Dates of onset not reported
Xercavins M, Llovet T, Navarro F, Morera MA, Moré J, Bella F, et al. Epidemiology of an unusually prolonged outbreak of typhoid fever in Terrassa, Spain. Clin Infect Dis. 1997 Mar;24(3):506–10.	Non-point source outbreak.
Yan M, Li X, Liao Q, Li F, Zhang J, Kan B. The emergence and outbreak of multidrug- resistant typhoid fever in China. Emerg Microbes Infect. 2016 Jun 22;5:e62.	Non-point source outbreak and no incubation period reported
Yang HH, Kilgore PE, Yang LH, Park JK, Pan YF, Kim Y, et al. An outbreak of typhoid fever, Xing-An County, People's Republic of China, 1999: estimation of the field effectiveness of Vi polysaccharide typhoid vaccine. J Infect Dis. 2001 Jun 15;183(12):1775–80.	No exposure identified

**APPENDIX 11: DETAILS OF STUDIES INCLUDED AND DATA
EXTRACTED IN THE SYSTEMATIC REVIEW AND META-
REGRESSION OF THE INCUBATION PERIOD OF *SALMONELLA*
TYPHI**

Please see accompanying USB drive

APPENDIX 12: CULLEN AND FREY PLOT SHOWING POTENTIAL DISTRIBUTION OF DATA FOR THE INCUBATION PERIOD OF *SALMONELLA* TYPHI

