

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Epidemiological investigations of surveillance strategies for zoonotic *Salmonella*

A dissertation presented
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
at Massey University

Jacqueline Benschop
2009

**Epidemiological investigations of surveillance
strategies for zoonotic *Salmonella***

A dissertation presented
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
at Massey University
Jacqueline Benschop

Institute of Veterinary, Animal and Biomedical Sciences
Massey University
Palmerston North, New Zealand

2009

(Submitted 16 January 2009)

Institute of Veterinary, Animal and Biomedical Sciences

Massey University

Palmerston North, New Zealand

2009

Abstract

This thesis is concerned with the application of recently developed epidemiological and statistical tools to inform the optimisation of a national surveillance strategy of considerable importance to human health. The results of a series of epidemiological investigations of surveillance strategies for zoonotic *Salmonella* are presented. *Salmonella* are one of the most common and serious zoonotic foodborne pathogenic bacteria globally. These studies were motivated by the increasing focus on the cost-effectiveness of surveillance while maintaining consumer confidence in food supply. Although data from the Danish *Salmonella* surveillance and control programme has been used in these investigations, the techniques may be readily applied to other surveillance data of similar quality.

The first study describes the spatial epidemiological features of Danish *Salmonella* surveillance and control programme data from 1995 to 2004, using a novel method of spatially adaptive smoothing. The conditional probability of a farm being a case was consistently high in the the south-west of Sonderjylland on the Jutland peninsula, identifying this area for further investigation and targeted surveillance. The identification of clustering of case farms led into the next study, which closely examines one year of data, 2003, for patterns of spatial dependency. *K*-function analyses provided evidence for aggregation of *Salmonella* case farms over that of all farms at distances of up to six kilometres. Visual semivariogram analyses of random farm-level effects from a Bayesian logistic regression model (adjusted for herd size) of *Salmonella* seropositivity, revealed spatial dependency between pairs of farms up to a distance of four kilometres apart. The strength of the spatial dependency was positively associated with slaughter pig farm density. We describe how this might inform the surveillance programme by potentially targeting herds within a four kilometre radius of those with high levels of *Salmonella* infection.

In the third study, farm location details, routinely recorded surveillance information, and industry survey data from 1995 were combined to build a logistic seroprevalence model.

This identified wet-feeding and specific pathogen free herd health status as protective factors for *Salmonella* seropositivity, while purchasing feed was a risk factor. Once adjusting for these covariates, we identified pockets of unexplained risk for *Salmonella* seropositivity and found spatial dependency at distances of up to six km (95% CI: 2–35 km) between farms. A generalised linear spatial model was fitted to the Jutland data allowing formal estimation of the range of spatial correlation and a measure of the uncertainty about it. There was a large within-farm component to the variance, suggesting that gathering more farm level information would be advantageous if this approach was to be used to target surveillance strategy.

The fourth study again considers data from the whole study period, 1995 to 2004. A detailed temporal analysis of the data revealed there was no consistent seasonal pattern and correspondingly no benefit in targeting sampling to particular times of the year. Spatio-temporal analyses suggested a local epidemic of increased seroprevalence occurred in west Jutland in late 2000. Lorelogram analyses showed a defined period of statistically significant temporal dependency, suggesting that there is little value in sampling more frequently than every 10 weeks on the average farm.

The final study uses findings from the preceding chapters to develop a zero-inflated binomial model which predicts which farms are most at risk of *Salmonella*, and then preferentially samples these high-risk farms. This type of modelling allows assessment of similarities and differences between factors that affect herd infection status (introduction) and those that affect the seroprevalence in infected herds (persistence and spread). The model suggested that many of the herds where *Salmonella* was not detected were infected but at a low prevalence. Using cost and sensitivity, we compared the results with those under the standard sampling scheme based on herd size, and the recently introduced risk-based approach. Model based results were less sensitive, but showed significant cost savings. Further model refinements, sampling schemes, and the methods to evaluate their performance are important areas for future work, and should continue to occur in direct consultation with Danish authorities.

Acknowledgements

Thank you to my three kind PhD supervisors who have provided guidance during this project. To Nigel French, who has mentored me and taught me about science and about myself. You have opened doors to many opportunities and been very influential in my career development. Thank you Nigel for encouraging me to publish and for always having a suggestion on how to move forward. To Mark Stevenson, who has mentored me and taught me a lot about epidemiology and the beauty there is in presenting data on a map. You have been very influential in my approach to teaching, learning, and thinking through a problem. Thank you Mark for your attention to detail and your tremendous ability to deliver complex information clearly. To Roger Morris, who invited me on board here at the EpiCentre and who made so many things possible for me, thank you.

Thank you to Martin Hazelton from the Statistics department for working alongside me on Chapter 3 of this thesis. Thanks to Simon Spencer for the same with Chapter 7. You both have shown me what can result from collaborating across disciplines and you have rekindled the interest I had in numbers.

Thanks to the people I have come to know at the Danish Meat Association. To Jan Dahl for provision of the data and advice on things Danish, I especially appreciated your hospitality while I was in Denmark. More recently Lis Alban has been my contact with the DMA. Jan and Lis, you both made it possible for me to work with data from the Danish *Salmonella* surveillance and control programme while I was in on the other side of the world. You provided thoughtful and frank comments on my work and have always been ready to consider new ideas. Also I thank Bodil Ydesen for providing additional risk factor and spatial data for this project.

For many friendly discussions and much helpful advice about modelling, spatial epidemiology, thesis writing, database management and the use of \LaTeX , WinBUGS, and R, I thank my workmates and fellow students Caryl Lockhart, Daan Vink, Naomi Cogger,

Eve Pleydell, Thibaud Porphyre, Birgit Schauer, Solis Norton, Patricia Jaros, and Petra Müllner.

Thank you Cord Heuer. Working alongside you I have been able to develop interest in and experience with other projects while doing this PhD. Cord, you have always been ready to give support and advice and to take back the reins to free me up to do this work. You, Naomi Cogger, Mark Stevenson, Eric Neumann, and Deb Prattley, have taken up some of my share of teaching allowing me to complete my PhD. Thank you.

Thanks to Colleen Blair, Julie Dunlop, Simon Vershaffelt, Christine Cunningham, and Wendy Maharey for your administrative and computing support. Thanks to Ruth Upper-ton for proof-reading the final drafts of this thesis.

I thank my other fellow students for making the EpiCentre a diverse and fascinating place to work. Too often here I have had to say good bye as someone returns home to Zambia, Bhutan, Uruguay, or elsewhere overseas to take up a senior role in animal health in their home country. That is both a rewarding and a sad part of the job.

I leave the most important to the end now and thank those who are very dear to me. To my husband for making our home a warm and loving place to return to at the end of each working day, for taking care of me and our children, and for cherishing the differences between us. There is no way I could have all the blessings that I do, and have completed this work, without you, Tim. And also to our children, Ruth, Tess, Ben and Katrina. I am so proud of you and love you very much. This is just the beginning.

Nomenclature

ACF	Autocorrelation function
AFP	Acute Flaccid Paralysis
ANN	Artificial neural networks
ARIMA	Auto-regressive integrated moving average
BOSS	Bovine Syndromic Surveillance System
BSE	Bovine spongiform encephalopathy
CCF	Cross-correlation function
CDC	Centers for Disease Control and Prevention
CHR	Central Husbandry Register
CI	confidence/credible interval
CJD	Creutzfeldt-Jakob Disease
DLM	Dynamic linear models
DMA	Danish Meat Association
DRB	Danish Meat Association risk-based sampling
DSSCP	Danish swine <i>Salmonella</i> surveillance and control programme
EARS	Early Aberration Reporting System
ED	Emergency department
ESR	Institute of Environmental Science and Research Limited
FMD	Foot and Mouth Disease
FoodNET	Foodborne Disease Surveillance Network
GAM	Generalised additive model
GIS	Geographic information system
GLM	Generalised linear model
HFRS	Haemorrhagic fever with renal syndrome
HIV	Human immunodeficiency virus

INAR	Integer-valued autoregressive models
MCMC	Markov Chain Monte Carlo
MRBA	Model derived risk-based sampling A
MRBB	Model derived risk-based sampling B
NNDSS	National Notifiable Disease Surveillance System
OD	Optical Density
OHS	Original herd size based sampling
OIE	World Organisation for Animal Health
RADAR	Rapid Analysis and Detection of Animal related Risk
RRV	Ross River virus
SARS	Severe Acute Respiratory Syndrome
SPF	specific pathogen free
SSI	Serological <i>Salmonella</i> Index
vCJD	variant Creutzfeldt-Jakob disease
WNV	West Nile Virus
ZIB	zero-inflated binomial

List of Publications

Benschop, J., Stevenson, M., Dahl, J., Morris R.S., French, N. (2009) Informing surveillance programmes by investigating spatial dependency of subclinical *Salmonella* infection. *Epidemiology and Infection* **137**:1348-1359

Benschop, J., Hazelton, M.L., Stevenson, M., Dahl, J., Morris, R.S., French, N. (2008) Descriptive spatial epidemiology of subclinical *Salmonella* infection in finisher pig herds: application of a novel method of spatially adaptive smoothing. *Veterinary Research* **39:02**

Benschop, J., Stevenson, M., Dahl, J., Morris R.S., French, N. (2008) Temporal and longitudinal analysis of Danish swine *Salmonella* control programme data: implications for surveillance. *Epidemiology and Infection* **136**:1511-1520

Benschop, J., Stevenson, M., Dahl, J., French, N. (2008) Towards incorporating spatial risk analysis for *Salmonella* seropositivity into the Danish swine surveillance programme. *Preventive Veterinary Medicine* **83**:347-359

Benschop, J., Hazelton, M., Stevenson, M., Dahl, J., Morris, R., French, N. (2007). 'Application of a novel method of spatially adaptive smoothing' in *Proceedings of the GisVet Conference*, University of Copenhagen, Denmark.

Benschop, J., Stevenson, M., Dahl, J., Morris, R., French, N. (2007). 'Using temporal and longitudinal analyses of accumulated data to inform sampling strategy' in *Proceedings from a Veterinary Epidemiology Seminar*, VetLearn Foundation, Palmerston North. ISSN: 1176-7979.

Benschop, J., Stevenson, M., Dahl, J., Morris, R., French, N. (2006). 'Can spatial analysis be used to target swine *Salmonella* surveillance for public health in Denmark?' in *Proceedings of GeoHealth Conference 2006*, Ministry of Health, Nelson, New Zealand. ISBN 0-478-30096-4.

Benschop, J., Stevenson, M., Dahl, J., Morris, R., French, N. (2006). 'second-order Spatial Effects: Danish Swine *Salmonella* Control Program' in *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*, Cairns, Australia.

Benschop, J., Stevenson, M., Dahl, J., Morris, R., French, N. (2006). 'Descriptive spatio-temporal epidemiology of sub-clinical *Salmonella* infection in Danish finisher pigs' in *Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics*, Cairns, Australia.

Benschop, J., Stevenson, M., Dahl, J., Morris, R., French, N. (2006). 'Spatial and risk factor analyses of *Salmonella* seropositivity in Danish pigs herds' in *Proceedings of the Epidemiology and Animal Health Management Branch, Food Safety, Animal Welfare and Biosecurity Branch of the New Zealand Veterinary Association Conference*, Vet Learn Foundation, Palmerston North, NZ. NZVA-FAVA Conference, Auckland, New Zealand. ISBN/ISSN: 1176-7979.

Contents

Abstract	iii
Acknowledgements	v
Nomenclature	vii
List of Publications	ix
Preface	xxi
1 Introduction	1
1.1 Introduction	1
1.2 The Danish Swine <i>Salmonella</i> Surveillance and Control Programme . .	3
1.2.1 Results from the control programme	4
1.3 The structure of this thesis	5
2 Literature review	7
2.1 Introduction	7
2.2 Surveillance	11
2.2.1 Active approaches to surveillance	11
2.2.2 Passive approaches to surveillance	12
2.2.3 Syndromic surveillance	14
2.2.4 Sentinel surveillance	16
2.2.5 Risk-based surveillance	17

2.3	Temporal surveillance	18
2.3.1	Time series methods	19
2.3.2	Statistical process control	49
2.3.3	Neural networks	53
2.3.4	The Temporal Scan statistic	54
2.4	Spatio-temporal Surveillance	54
2.4.1	Spatial variation in risk	57
2.4.2	Spatial and spatio-temporal clustering	59
2.4.3	Other spatio-temporal surveillance techniques	66
2.5	Conclusions	68
3	Descriptive spatial analysis of <i>Salmonella</i> infection in Danish pig herds	69
3.1	Abstract	69
3.2	Introduction	70
3.3	Materials and methods	71
3.3.1	The <i>Salmonella</i> Surveillance and Control Programme	71
3.3.2	The data	72
3.3.3	Statistical analyses	73
3.4	Results	76
3.4.1	Summary statistics	76
3.4.2	Spatial analysis	76
3.5	Discussion	83
4	Investigation of spatial dependency to inform surveillance	87
4.1	Abstract	87
4.2	Introduction	88
4.3	Materials and methods	90
4.3.1	The data set	90

	xiii
4.3.2	Pig-level data 91
4.3.3	Farm-level data 91
4.3.4	Spatial analysis 91
4.4	Results 94
4.5	Discussion 106
4.6	Acknowledgement 110
5	Risk factor and spatial analysis 111
5.1	Abstract 111
5.2	Introduction 112
5.3	Materials and methods 114
5.3.1	Data description and handling 114
5.3.2	Risk factor analysis 114
5.3.3	Spatial analysis 116
5.4	Results 118
5.5	Discussion 126
5.6	Conclusion 129
5.7	Acknowledgements 130
6	Temporal and Longitudinal analysis 131
6.1	Abstract 131
6.2	Introduction 132
6.3	Materials and methods 133
6.3.1	The Danish swine <i>Salmonella</i> surveillance and control programme 133
6.3.2	The data 134
6.3.3	Statistical analysis 134
6.4	Results 137
6.5	Discussion 148
6.6	Acknowledgements 152

7	Predictive modelling of herd-level prevalence for risk-based surveillance	153
7.1	Abstract	153
7.2	Introduction	154
7.3	Materials and methods	156
7.3.1	Data sources	156
7.3.2	Sampling schemes	157
7.3.3	Model development for the sampling schemes	159
7.3.4	Comparison of the sampling schemes	162
7.4	Results	163
7.4.1	Data sources	163
7.4.2	Model development for the sampling schemes	163
7.4.3	Comparison of the sampling schemes	166
7.5	Discussion	176
7.5.1	A discussion of sampling	178
7.5.2	A discussion of future work	179
7.5.3	A discussion of bias, confounding, and chance	180
8	General Discussion	181
8.1	Introduction	181
8.1.1	From on-farm to slaughterhouse interventions	182
8.1.2	The change in human cases of salmonellosis in Denmark	183
8.2	Lessons learnt	184
8.2.1	The value of multi-disciplinary collaboration	184
8.2.2	The importance of data quality	185
8.3	Future perspectives	186
8.3.1	Future work for these data	186
8.3.2	Is risk-based sampling ‘safe’?	188

8.3.3	Continual improvement of visualisation of surveillance data . . .	188
8.3.4	Innovative surveillance	189
8.4	Conclusion	192
A	Appendix 1	A-232
A.1	Introduction	A-232
A.2	Materials and methods	A-232
A.2.1	Data description and handling	A-232
A.2.2	Risk factor analysis	A-232
A.3	Results	A-233
A.4	Discussion	A-235

List of Figures

2.1	<i>Salmonella enteritidis</i> , Netherlands, 2002–2004.	9
2.2	Surveillance pyramid for case ascertainment.	13
2.3	Useful data sources for syndromic surveillance.	15
2.4	Location of arbovirus sentinel cattle herds in New Zealand, 2008.	19
2.5	Respiratory deaths data, raw monthly time series, UK, 1974–1979.	21
2.6	Nosocomial infections in a Spanish hospital, 1982–1990.	22
2.7	Weekly <i>Salmonella typhimurium</i> cases, Denmark, 2005–2008.	23
2.8	Respiratory deaths data, loess smoothed monthly time series, UK, 1974– 1979.	25
2.9	Respiratory deaths data, detrended monthly time series, UK, 1974–1979.	27
2.10	Respiratory deaths data, monthly box plot of time series, UK, 1974–1979.	30
2.11	Respiratory deaths data, decomposition of time series, UK, 1974–1979.	31
2.12	Respiratory deaths data, lagged scatterplots, UK, 1974–1979.	33
2.13	Respiratory deaths data, autocorrelation function plot, UK, 1974–1979	34
2.14	Respiratory deaths data, raw and smoothed periodograms, UK, 1974–1979	47
2.15	Campylobacteriosis in New Zealand, 2006–2007.	52
2.16	Point map of Broad Street cholera cases, London, 1854.	58
3.1	Map of Denmark showing location of counties.	79
3.2	Kernel smoothed maps of Danish pig herd densities, 1995–2004.	80
3.3	Kernel smoothed maps of Danish pig densities, 1995–2004.	81
3.4	Kernel smoothed maps of conditional probabilities, 1995–2004.	82

4.1	Map of Denmark showing areas used in the K -function analysis.	99
4.2	Kernel smoothed maps of Danish pig herd densities, 2003.	100
4.3	Inhomogeneous K -function for farms in north Jutland, 2003.	101
4.4	Boxplot of nearest neighbour distances for four Danish counties, 2003. .	102
4.5	Observed difference K -function, Denmark, 2003.	103
4.6	Semivariograms, major pig-producing Danish counties, 2003.	104
4.7	Semivariograms, minor pig-producing Danish counties, 2003.	105
5.1	Map of Denmark showing study herd locations, 1995.	123
5.2	Intensity plot of the random farm effects, Denmark, 1995.	124
5.3	Spatial semivariograms fitted to random farm effects, Denmark, 1995. .	125
6.1	Time series plots stratified by positivity, Denmark, 1995–2004.	140
6.2	Time series plots stratified by region, high positive strata, Denmark, 1995– 2004.	141
6.3	Time series plots stratified by region, Denmark, 1995–2004.	142
6.4	Time series residual plots stratified by region, Denmark, 1995–2004. . .	143
6.5	Monthplot of the time series residuals, Denmark, 1995–2004.	144
6.6	Periodograms of the time series residuals, Denmark, 1995–2004.	145
6.7	Plot of forecasted time series, Denmark, 1995–2004.	146
6.8	Stratified lorelograms from DSSCP data, Denmark, 2002–2004.	147
7.1	Frequency histograms, number of pigs sampled, Denmark, 2003 and 2004	171
7.2	Frequency histograms, within-herd seroprevalence, Denmark, 2003 and 2004	172
7.3	Scatter plot of seroprevalence vs. infection probability, Denmark, 2003.	173
7.4	Scatter plot, random farm effects A_i , 2003 vs. 2004, Denmark.	174
7.5	Scatter plot, random farm effects B_i , 2003 vs. 2004, Denmark.	175

List of Tables

3.1	<i>Salmonella</i> seropositivity, stratified by county, Denmark, 1995–2004. . .	78
4.1	Descriptive statistics, areas for <i>K</i> -function analysis, Denmark, 2003. . .	97
4.2	Logistic regression model output, Denmark, 2003.	98
5.1	Descriptive statistics for 3784 Danish finisher pig herds, 1995.	120
5.2	Informed priors used for fixed effects in logistic regression model. . . .	121
5.3	Logistic regression model output, complete cases, Denmark, 1995. . . .	122
7.1	Logistic regression model output, Denmark, 2003.	167
7.2	Zero-inflated model output: infection status, Denmark, 2003.	168
7.3	Zero-inflated model output: seropositivity, Denmark, 2003.	169
7.4	Performance comparison of four sampling schemes.	170
A.1	Logistic regression model output, with missing values imputed, Denmark, 1995.	A-234

Preface

Family Story

'Poor little shitter,' her sister said. A tainted wonton was what did it, that third night of the power cut. Lines down everywhere, and us out in the sticks,

The power company put us in a priority queue and sick of it all we hit town and queued instead at the noodle-house - fuggy, crowded with pale

parents and fractious children also escaping their darkened homes and wanting light, light and warm food. The rain steamed off our backs

An ease came as we waited our turn, and children played among tables. A kind of company, that's what we were, such as Chaucer or Boccaccio made a meal of,

only we made a meal of numbers on the menu, our children too unsure to pronounce the names - except Katrina, whose seven-year-old mouth

had eaten many wontons and called them out now with conviction. On the wall, the Health Inspection rating was unreadable behind yellowing

cellophane, a detail recalled as, hours later in the cold bedroom, we stroked Katrina's clammy forehead while she writhed. . . . Poor little shitter, asleep now,

already a family story, her stained sheets churning
away in a washing machine that had at last,
when our hope was gone, our patience spent,

at the very last - then it was, as tiny red standby
lights throughout the house glowed into life,
the blessed machine shook, stirred itself and beeped.

Tim Upperton

from 'A House on Fire' Steele Roberts (2009)